

**Test-retest reliability of the SSQ-12 questionnaire:  
For hearing-aid wearers using pen-and-paper  
administration method**

---

A thesis submitted in partial fulfilment of the requirements for the degree  
of

**Master of Audiology**

in the Department of Communication Disorders

at the University of Canterbury

---

Bethany Cox

*University of Canterbury*

2016

## Acknowledgements

To my family – thank you Mum and Dad for setting an example of work ethic and values that has helped me reach this point, and for your constant love and encouragement. Thank you to my brothers (Sam, Jonny, and Tim) for being there and encouraging me along the way. Thank you to my sisters – Elya for your practical support and listening ear, and for doing things like running a hot bath when the pressure was high but work efficiency low! Thank you Melody and Tim for letting me hijack my niece and nephews for fun days out when I needed a break, and paying me in study snacks for babysitting. Thank you Grandad for your constant prayers, encouragement and general interest in my studies. Thank you Poppa for being a guinea-pig in my audiology assignments, including this thesis!

To my flatmate – thank you Rachal for your support and interest in my studies, and your patience when I lapsed in household chores, or spent weeks with my computer camped out on the dining table! And thank you Tyke (the dog) for keeping me company at home and forcing me to take long walks.

To my classmates – I couldn't have asked for a better bunch of people to study with. Thank you for your support along the way, whether it be helping me understand statistics, reassuring me when the pressure was on, or just being fun and encouraging to be around. A particular thanks to Chae, my partner in crime for thesis-izing! It's been awesome to work on a thesis project with you – I appreciate your optimistic and fun manner.

To my supervisor – thank you Rebecca for your legendary support! I am incredibly grateful for the privilege of doing my thesis under your supervision, given

your expertise and reputation of being an all-round great and reliable supervisor!

Thank you for your effort, timely and constructive feedback, holding me accountable, and giving me a ‘hoozle-up’ when I needed it!

Lastly, I want to acknowledge God, who is the giver of life and has guided me to this point. My prayer is that I would continue to use my skills for his glory and ultimate purposes.

## **Abstract**

The aim of this study was to determine if the SSQ-12 (Speech, Spatial, and Qualities of Hearing Scale – Short form) is a reliable questionnaire to assess hearing-aid benefit for experienced hearing-aid wearers using the pen-and-paper administration method. Twenty-eight experienced hearing-aid wearers were recruited from the University of Canterbury's audiology clinic database and from the general public. Participants were sent the SSQ-12 questionnaire 3 times at 6-week intervals. The participants' responses across the three different administration times (T0, T1, and T2) were compared using a repeated measures ANOVA to determine if their answers remained stable over time, when no intervention was occurring. The results showed there were no significant differences between the SSQ-12 total or sub-scale scores for each participant's T0, T1, and T2 data. Critical change scores were calculated for total, and sub-scale scores, to facilitate clinicians identifying whether a change in score is clinically significant. In conclusion, the results of this study indicate the SSQ-12 has good test-retest reliability for experienced hearing-aid wearers using the pen-and-paper administration method.



# Table of Contents

<b>Test-retest reliability of the SSQ-12 questionnaire: For hearing-aid wearers using pen-and-paper administration method .....</b>	<b>i</b>
<b>Acknowledgements .....</b>	<b>ii</b>
<b>Abstract .....</b>	<b>iv</b>
<b>List of Abbreviations .....</b>	<b>viii</b>
<b>List of Figures .....</b>	<b>x</b>
<b>List of Tables .....</b>	<b>xi</b>
<b>Chapter One: Review of Literature .....</b>	<b>1</b>
<b>1.1 Overview .....</b>	<b>1</b>
1.1.1 Prevalence and impact of hearing loss. ....	1
1.1.2 The ear and hearing. ....	2
1.1.3 Hearing loss. ....	3
1.1.4 International classification of functioning, disability, and health (ICF).....	4
<b>1.2 Outcome Measures in Audiology .....</b>	<b>6</b>
1.2.1 Overview .....	6
1.2.2 Objective outcome measures. ....	7
1.2.3 Subjective outcome measures.....	7
<b>1.3 Self report measures .....</b>	<b>8</b>
1.3.1 Overview. ....	8
1.3.2 Brief history of self-report measures in audiology. ....	9
1.3.3 Self-report measures in audiology.....	10
1.3.4 Variations of self-report measures.....	11
1.3.5 Specific populations. ....	12
1.3.6 Open-end versus closed-end questions and response format. ....	12
1.3.7 Limitations of self-report measures.....	14
1.3.8 Value of self-report measures.....	16
<b>1.4 Psychometric Issues.....</b>	<b>18</b>
1.4.1 Overview. ....	18
1.4.2 Validity.....	18
1.4.3 Reliability.....	21
1.4.4 Normative data.....	27
<b>1.5 Original SSQ-49 (Speech, Spatial, and Qualities of Hearing Scale) .....</b>	<b>27</b>
1.5.1 Rationale for the SSQ-49.....	27
1.5.2 Auditory scene analysis.....	29
1.5.3 Description of the SSQ-49.....	30
1.5.4 Development of the SSQ-49.....	31
1.5.5 Hearing Handicap Questionnaire (HHQ). ....	33
1.5.6 Primary SSQ-49 studies. ....	34
1.5.7 Research using the SSQ-49.....	35
1.5.8 SSQ covariates.....	38
1.5.9 Variations of the original SSQ-49.....	41
<b>1.6 Psychometric properties of the SSQ-49 .....</b>	<b>44</b>
1.6.1 Validity of the SSQ-49.....	44
1.6.2 Reliability of the SSQ-49.....	59

<b>1.7 SSQ-49 Normative data .....</b>	<b>63</b>
<b>1.8 SSQ-12 .....</b>	<b>67</b>
1.8.1 Introduction. ....	67
1.8.2 Description of the SSQ-12.....	67
1.8.3 Development of the SSQ-12.....	68
1.8.4 Research using the SSQ-12.....	68
1.8.5 Psychometric properties of the SSQ-12.....	69
1.8.6 SSQ-12 Normative Data.....	70
<b>1.9 Gap in the literature .....</b>	<b>71</b>
<b>1.10 Research question .....</b>	<b>71</b>
<b>1.11 Hypotheses.....</b>	<b>71</b>
<b>Chapter Two: Methods .....</b>	<b>72</b>
<b>2.1 <i>A priori</i> power analysis.....</b>	<b>72</b>
<b>2.2 Ethics approval .....</b>	<b>72</b>
<b>2.3 Participants .....</b>	<b>72</b>
2.3.2 Recruitment. ....	72
2.3.3 Inclusion/exclusion criteria.....	73
2.3.4 Drop-outs.....	73
<b>2.4 Measures.....</b>	<b>73</b>
2.4.1 Pure-tone audiometry. ....	73
2.4.2 Demographic questionnaire.....	74
2.4.3 Hearing Handicap Questionnaire (HHQ). ....	74
2.4.4 Speech, Spatial and Qualities of Hearing Scale – short version (SSQ-12). ....	74
<b>2.5 Procedure .....</b>	<b>75</b>
<b>2.6 Statistical Analyses .....</b>	<b>76</b>
<b>Chapter Three: Results .....</b>	<b>77</b>
<b>3.1 Overview.....</b>	<b>77</b>
<b>3.2 Participant description.....</b>	<b>77</b>
<b>3.3 Correlations between SSQ-12 scores and participant characteristics</b> <b>(covariates) .....</b>	<b>80</b>
<b>3.4 Test-retest reliability .....</b>	<b>80</b>
<b>3.5 Confidence intervals and critical difference scores.....</b>	<b>83</b>
<b>Chapter 4: Discussion.....</b>	<b>84</b>
<b>4.1 Research question .....</b>	<b>84</b>
<b>4.2 Comparison with the literature .....</b>	<b>84</b>
4.2.1 Comparison with test-retest reliability data from other hearing related questionnaires. ....	84
4.2.2 Comparison with test-retest reliability data from the SSQ-49. ....	86
<b>4.3 Critical difference scores.....</b>	<b>88</b>
4.3.1 Comparison with critical difference scores from other hearing related questionnaires. ....	88
4.3.2 Comparison with critical difference scores from the SSQ-49.....	88
<b>4.4 Research limitations .....</b>	<b>89</b>
<b>4.5 Future research.....</b>	<b>90</b>
<b>4.6 Clinical implications .....</b>	<b>91</b>
<b>4.7 Summary .....</b>	<b>91</b>
<b>References.....</b>	<b>93</b>
<b>Appendices .....</b>	<b>112</b>

<b>Appendix A: Ethics approval letter. ....</b>	<b>112</b>
<b>Appendix B: Poster advertisement. ....</b>	<b>113</b>
<b>Appendix C: Research information and consent forms.....</b>	<b>114</b>
<b>Appendix D: Demographic questionnaire.....</b>	<b>118</b>

## **List of Abbreviations**

APHAB	– Abbreviated Profile of Hearing Aid Benefit
ANOVA	– analysis of variance
BEPTA	– better ear pure-tone average
CHL	– conductive hearing loss
CI	– cochlear implant
COSI	– Client Oriented Scale of Improvement
CPHI	– Communication Profile for the Hearing Impaired
EAC	– external auditory canal
GHABP	– Glasgow Hearing Aid Benefit Profile
HA	– hearing aid
HAPI	– Hearing Aid Performance Inventory
HHIE/A	– Handicap Inventory for the Elderly/Adults
HHS	– Hearing Handicap Scale
HHQ	– Hearing Handicap Questionnaire
IHC	– inner hair cell
IOI-HA	– International Outcome Inventory for Hearing Aids
L.I.F.E.	– Listening Inventory for Education
OHC	– outer hair cell
ONH	– older normal hearing adults
NCIQ	– The Nijmegen Cochlear Implant Questionnaire
PEACH	– Parents Evaluation of Aural/Oral Performance of Children
PHAB	– Profile of Hearing Aid Benefit
PTA	– pure-tone average
REAG	– real-ear aided gain
REIG	– real-ear insertion gain
SADL	– Satisfaction with Amplification in Daily Life
SHQ	– Spatial Hearing Questionnaire

SNHL – sensorineural hearing loss

SOAC – Significant Other Assessment of Communication

SSQ – Speech Spatial and Qualities of Hearing Scale

SSQ-B – Speech Spatial and Qualities of Hearing Scale – benefit version

SSQ-C – Speech Spatial and Qualities of Hearing Scale – comparison version

SSQ-Children – Speech Spatial and Qualities of Hearing Scale – child version

SSQ-Parent – Speech Spatial and Qualities of Hearing Scale – parent version

SSQ-Teacher – Speech Spatial and Qualities of Hearing Scale – teacher version

SSQ-5 – Speech Spatial and Qualities of Hearing Scale – screening version

SSQ-12 – Speech Spatial and Qualities of Hearing Scale – short version

SSQ-15 – Speech Spatial and Qualities of Hearing Scale – German short version

SSQ-49 – Speech Spatial and Qualities of Hearing scale – original version

TEACH – Teacher’s Evaluation of Aural/Oral Performance of Children

TM – tympanic membrane

WEPTA – worse ear pure-tone average

WHO – World Health Organization

YNH – younger normal hearing adults

## List of Figures

<i>Figure 1:</i> Graphic representation of the 2001 International Classification of Functioning, Disability, and Health (ICF) framework (World Health Organization, 2002). .....	5
<i>Figure 2:</i> Examples of response scales in the SSQ-49 (MRC Institute of Hearing Research, n.d.). .....	31
<i>Figure 3:</i> Mean audiogram of participants for right and left ears, with standard error of measurement indicated by the vertical bars (N = 28). .....	78

## List of Tables

Table 1: Description of the types of research involving the SSQ-49 and references to the studies that have used the SSQ-49.....	36
Table 2: Studies showing a relationship between the SSQ-49 and localisation tests (significant correlations are in bold).....	48
Table 3: Studies showing correlations between the SSQ-49 and other hearing-related self-report measures (handicap and disability).....	50
Table 4: Studies reporting on correlations between SSQ-49 scores and speech perception tests. ....	53
Table 5: Studies reporting on the relationship between SSQ-49 scores and hearing ability (PTA).....	57
Table 6: Test-retest correlations of the SSQ-49 (total and sub-scale scores) for four different combinations of administration methods (n = 39 for the interview – mail group, and n = 40 for the remaining groups). Adapted from Singh and Pichora-Fuller (2010). ....	60
Table 7: Test-retest reliability data from hearing-related self-report measures for individuals with hearing loss (except some in the SSQ study who had normal hearing) using the pen-and-paper administration method. ....	62
Table 8: Normative SSQ-49 data from Banh et al., 2012; Gatehouse & Noble, 2004; Demeester et al., 2012; and Moulin & Richard, 2016. ....	66
Table 9: Normative SSQ-49 data from Singh & Pichora-Fuller, 2010. ....	67
Table 10: Mean, standard deviation, and range of days between administration times. ....	76
Table 11: Mean, standard deviation and range values for participants' age, self-rated severity of hearing loss, HHQ scores, BEPTA, WEPTA, and years of HA use (N = 28).....	77
Table 12: Sample percentages of ethnicity, income, and level of education (N = 28). ....	78
Table 13: Minimum, maximum, mean and standard deviation of the SSQ-12 scores for total and sub-scale scores from T0, T1, and T2 data (N = 28). ....	79
Table 14: Correlation matrix for self-rated severity of hearing loss, BEPTA, WEPTA, and mean HHQ scores (N = 28). ....	79
Table 15: Pearson correlation coefficients of mean SSQ-12 scores, and possible covariates (age, self-rated severity of hearing loss, BEPTA, WEPTA, and mean HHQ scores). Significant correlations are indicated in bold. ....	80

Table 16: Pearson's correlation matrix of T0, T1, and T2 data for total SSQ-12 scores, and speech, spatial, and qualities sub-scale scores. The correlations in red bold indicate the test-retest reliability of the SSQ-12 using Pearson correlation values for the total, speech, spatial, and qualities sections across administration times T0-T1, T0-T2, and T1-T2.....	82
Table 17: Lower and upper limits of the 95% confidence intervals and critical difference scores for total and sub-scale scores from T0, T1, and T2 data. ....	83



# Chapter One: Review of Literature

## 1.1 Overview

### 1.1.1 Prevalence and impact of hearing loss.

Disabling hearing loss is defined as hearing loss greater than 40 dB HL in the better ear for adults, and greater than 30 dB HL for children (World Health Organization, 2016). According to the World Health Organization (2016), hearing loss affects 360 million people worldwide or 5.3% of the world's population. This figure includes 328 million adults and 32 million children (World Health Organization, 2016).

Hearing loss restricts the ability to communicate optimally using spoken language. If someone cannot hear properly, it affects their ability to learn spoken language, listen to verbal dialogue, and interact conversationally. Hearing loss leads to communication breakdowns, because the person may not hear and respond appropriately to speech or conversation. Consequences for the individual include the functional, emotional and psychosocial domains (World Health Organization, 2015).

In children, hearing loss can affect speech and oral language development, and influence social and emotional development, and academic progress (Flexer & Madell, 2014). In older populations, hearing loss has been linked to perceptions of reduced quality of life (Ciorba, Bianchini, Pelucchi, & Pastore, 2012; Dalton et al., 2003), and feelings of social isolation, loneliness, depression, and anxiety (Heine & Browning, 2002; Weinstein, 2015; Weinstein & Ventry, 1982). Other emotions experienced may include embarrassment, shame, apathy, and frustration (Tye-Murray, 2015). A reduced ability to communicate can also lead to strained

relationships between the person with hearing loss and their frequent communication partners (Kamil & Lin, 2015). In addition, for working adults hearing loss may impact on employment opportunities and their ability to work as effectively as co-workers with normal hearing (Emmett & Francis, 2014; Winn, 2007).

### **1.1.2 The ear and hearing.**

The peripheral auditory system includes the outer, middle, and inner ear components and the auditory nerve (Musiek & Baran, 2007). The outer ear consists of the pinna and external auditory canal (EAC). The pinna collects and funnels sound waves down the EAC to the tympanic membrane (TM). The middle ear is made up of the TM, middle ear space, and ossicles: malleus, incus and stapes (Musiek & Baran, 2007). The middle ear is connected to the eustachian tube, which equalises pressure in the middle ear space with atmospheric pressure (Musiek & Baran, 2007). The ossicles form the ossicular chain, which conducts sound vibrations from the TM to the oval window at the cochlea. The inner ear consists of the cochlea, which is connected to the vestibular organ. Within the cochlea, is the basilar membrane. The stapes in the middle ear moves in and out of the oval window, disturbing the cochlear fluids, and setting up a travelling wave on the basilar membrane which propagates from the base of the basilar membrane (high frequencies) up to the apical end (low frequencies) (Musiek & Baran, 2007). Along the length of the basilar membrane is the organ of Corti, which houses the outer hair cells (OHCs) and inner hair cells (IHCs). The OHCs amplify soft sounds and the IHCs translate the mechanical vibrations on the basilar membrane into an electrical signal that travels up the auditory nerve (8<sup>th</sup> cranial nerve) to the auditory cortex (Musiek & Baran, 2007). Abnormalities in the structure or function of part of the peripheral or central auditory systems can result in hearing loss. Dysfunction may be due to genetic causes (e.g., syndromic, non-

syndromic, hereditary) or environmental causes (e.g., trauma, infection). Dysfunction may affect both ears (bilateral hearing loss) or one ear (unilateral hearing loss).

### **1.1.3 Hearing loss.**

#### ***1.1.3.1 Conductive hearing loss.***

A conductive hearing loss (CHL) occurs when dysfunction occurs either in the outer or middle ear (Musiek & Baran, 2007). Causes of CHL include otitis media, ossicular discontinuity, aural atresia, otosclerosis, and cholesteatoma. CHL is evident from pure-tone audiometry, where bone-conduction thresholds are within normal limits, and there is an air-bone gap of 15 dB HL or more between the bone conduction thresholds and the air-conduction thresholds (Schlauch & Nelson, 2015). Conductive hearing losses may resolve over time or be corrected by surgery (Tye-Murray, 2015)

#### ***1.1.3.2 Sensorineural hearing loss.***

Sensorineural hearing loss (SNHL) occurs when there is dysfunction either in the inner ear or on the auditory nerve. Causes of SNHL include, presbycusis, noise-induced hearing loss, viral or bacterial infections affecting the inner ear, and structural abnormalities of the cochlea or auditory nerve. The pattern of loss on the audiogram for a SNHL shows similar bone-conduction and air-conduction thresholds ( $\leq 10$  dB HL air-bone gap; Schlauch & Nelson, 2015).

#### ***1.1.3.3 Mixed hearing loss.***

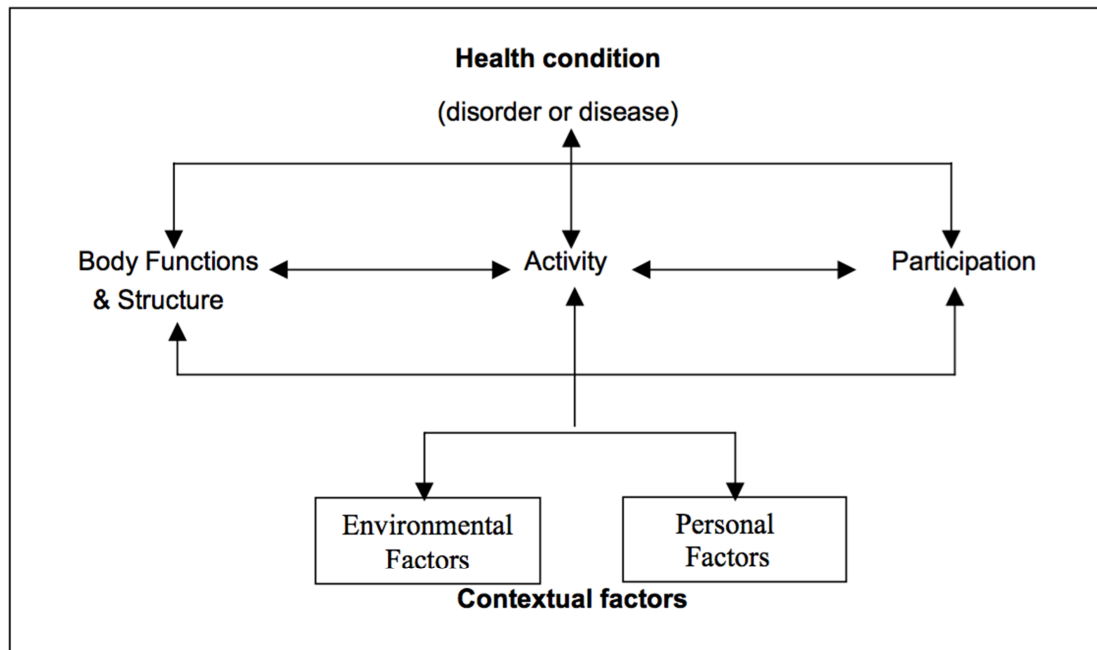
Hearing loss can be a combination of both CHL and SNHL, i.e., a mixed hearing loss. The pattern of loss on the pure-tone audiogram shows bone-conduction thresholds outside normal limits (indicating SNHL), combined with a significant air-bone gap ( $\geq 15$  dB HL) between bone-conduction and air-conduction thresholds (indicating some degree of CHL; Schlauch & Nelson, 2015).

#### **1.1.4 International classification of functioning, disability, and health (ICF).**

The World Health Organization's (WHO) International Classification of Functioning, Disability, and Health (ICF; 2001) is a framework for classifying the impact of impairment, subsequent activity and participation restrictions, and the personal and environmental contextual factors influencing an individual's ability to function normally. The 2001 version is a revision of the International Classification of Impairments, Disabilities, and Handicaps (World Health Organization, 1980). The revised model uses more positive terminology with a focus on what the person can do (i.e., functioning, health, activity and participation restrictions), rather than on what they cannot do (impairment, disability, and handicap; Stephens, 2001; World Health Organization, 2002). It also adds contextual factors (personal and environmental) and indicates two-way relationships between each of the factors (see Figure 1 below; Stephens, 2001).

In the current model, body functions and structure refer to the impairment at the body level, e.g., damaged outer hair cells in the cochlea. Activity limitations refer to "difficulties an individual may have in executing a task or action" (World Health Organization, 2002, p. 10), e.g., listening to and understanding speech in background noise. Participation restrictions are "problems an individual may experience in involvement in life situations" (World Health Organization, 2002, p. 10), e.g., conversing with colleagues in the staff room. Environmental factors include physical, social (legal and social structures), and attitudinal factors that can be facilitators or barriers to an individual's functioning (World Health Organization, 2002), e.g., a reverberant staff room (physical barrier). Personal factors include age, gender, coping

strategies, education, profession, character and social background that influence how a person experiences hearing loss (World Health Organization, 2002).



*Figure 1: Graphic representation of the 2001 International Classification of Functioning, Disability, and Health (ICF) framework (World Health Organization, 2002).*

#### **1.1.4.1 Disability and handicap.**

According to the WHO classification of impairments, disabilities, and handicaps (1980), disability refers to “any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner of within the range considered normal for a human being” (World Health Organization, 1980, p. 28). Handicap refers to “a disadvantage for a given individual, resulting from an impairment or disability, that limits or prevents the fulfilment of a role that is normal for that individual” (World Health Organization, 1980, p. 29). In this model, handicap is a social phenomenon, but also incorporates the emotional consequences of an auditory impairment and disability resulting from disturbed relationships, isolation, and alienation. A simplification of this model is to say that disability refers to the auditory consequences of hearing loss, i.e., not being able to hear softly spoken

people or hear in background noise, whereas handicap encompasses the non-auditory consequences of hearing loss, i.e., emotional distress and social restrictions as a result of hearing loss.

The WHO 1980 model using the terms impairment, disability and handicap will be used herein for the following reasons: (1) activity and participation restrictions do not provide a clear distinction between the two categories, because an activity limitation could be mistakenly categorised as a participation restriction, or vice versa (Noble, 2013; Stephens, 2001), (2) participation restriction does not immediately account for the emotional consequences of hearing loss (Noble, 2013; Stephens, 2001), (3) the authors of the SSQ-12 have opted to use the WHO 1980 definitions of impairment, disability, and handicap for the reasons mentioned above (Noble, Tyler, Dunn, & Bhullar, 2008). To provide consistency between the relevant research literature and this research, the WHO 1980 definitions of disability and handicap will be used in this report (as opposed to the later WHO, 2001 ICF model definitions: activity limitations, and participation restrictions, etc.).

## **1.2 Outcome Measures in Audiology**

### **1.2.1 Overview**

Outcome measures are assessments used to demonstrate the value or benefit of a particular treatment or intervention, and assess whether rehabilitation goals have been met (Cox, 2003). Outcome measures are necessary in order to justify a treatment or intervention by demonstrating (identifying and quantifying) a positive difference to an individual's before-treatment status or functional ability (Johnson & Danhauer, 2002; Weinstein, 2000). In audiology, an important part of demonstrating the efficacy of hearing aids is to help justify the cost of them, either for the individual purchasing

their own hearing aids, or for a third-party payer (Johnson & Danhauer, 2002).

Outcome measures can be objective or subjective.

### **1.2.2 Objective outcome measures.**

Objective outcome measures are typically carried out by the audiologist during appointments and include measuring hearing-aid characteristics and differences in performance between aided and unaided conditions. Typical objective outcome measures include real-ear measures (e.g., real-ear aided gain [REAG], and real-ear insertion gain [REIG]), the speech intelligibility index, speech recognition measures, and functional gain in the sound field (Cox, 2003; Johnson & Danhauer, 2002; Weinstein, 2000). Hearing-aid use via data logging can also be used as an objective outcome measure (Cox, 2003). The relationship between objective outcome measures mentioned above, and subjective outcome measures (described below) are not always clear. Therefore, it is becoming increasingly important in today's patient-centred health-care model to assess both objective and subjective outcome measures (Weinstein, 1997).

### **1.2.3 Subjective outcome measures.**

Subjective outcome measures provide information on the client's opinion of their treatment and outcomes in real-life. Typically, in audiology, subjective outcome measures are assessed formally using self-report inventories or questionnaires (although it may also be done more informally by open-ended questions and general discussion between the clinician and patient). Cox (2003) identified seven domains for measuring subjective hearing-aid outcomes; satisfaction, quality of life, benefit, use, impact on others (third-party disability), residual participation restrictions (i.e., handicap), and residual activity limitations (i.e., disability). Common audiology-based inventories that address one or more of the domains above include the Hearing

Handicap Inventory for the Elderly/Adults (HHIE/A; Weinstein & Ventry, 1982; Newman, Weinstein, Jacobson, & Hug, 1990), Communication Profile for the Hearing Impaired (CPHI; Demorest & Erdman, 1987), Abbreviated Profile of Hearing Aid Benefit (APHAB; Cox & Alexander, 1995), Satisfaction with Amplification in Daily Life (SADL; Cox & Alexander, 1999), Glasgow Hearing Aid Benefit Profile (GHABP; Gatehouse, 1999), Hearing Aid Performance Inventory (HAPI; .Walden, Demorest, & Hepler, 1984), and International Outcome Inventory for Hearing Aids (IOI-HA; Cox et al., 2000).

### **1.3 Self report measures**

#### **1.3.1 Overview.**

Self-report measures are questionnaires that probe something of interest to measure, such as disability or handicap related to hearing impairment. The person who is experiencing hearing difficulties completes the self-report measure. Therefore, self-report is by nature a subjective rather than an objective measure (Noble, 1998).

Any information collected from a patient's own telling of their experiences comes under the banner of self-report and is therefore subjective in nature, e.g., clinical interviews, or case histories (Noble, 1978). Self-report information can be assessed informally, e.g., in an unstructured interview or case history, or it can be assessed formally in a structured manner where they complete a questionnaire. A self-report measure that provides a quantifiable measure of a person's responses (e.g., HHIE/A) can more easily be compared over time and between different people (Weinstein, 1993).

Self-report measures are designed to measure a particular construct such as hearing handicap or disability. Typical self-report measures comprise a set of



questions related to the construct being measured and incorporate different scenarios. Patients choose from a set of response alternatives or place a mark on a fixed scale ranging from one viewpoint to an opposing viewpoint (Gatehouse & Akeroyd, 2007). More recently, less structured self-report measures such as the Client Oriented Scale of Improvement (COSI; Dillon, James, & Ginis, 1997) ask patients to nominate their own goals, and then rate their ability pre- and post-intervention.

### **1.3.2 Brief history of self-report measures in audiology.**

The earliest record of systematic self-report use in hearing services was in 1886, by an otologist named Barr (Noble, 1993). Barr recognised the effect hearing loss had on the “social comforts” of those affected and the further effects on the “social fabric” where they existed, and hence he recognised the importance of addressing not only their impairment (i.e., measuring their pure-tone thresholds) but their perceived disability and handicap as well (Noble, 1993, p. 299). Today, self-report measures are considered valuable tools in audiology for assessing the need for and efficacy of rehabilitation (American Speech-Language-Hearing Association, 2006; British Society of Audiology, 2012; Cox, 2005; New Zealand Audiological Society, 2015; Tye-Murray, 2015); but this was not always the case (Erdman, 1993; Noble, 2013).

There are a number of possible reasons why self-report measures were not routine in audiology assessments until more recently. Noble (2013) outlined the early controversy that surrounded self-report tools for measuring psychological phenomena. He stated that the subjective measures did not align with the scientific model that demanded empirical and objective data, and there was concern about the accuracy and reliability of such measures (Noble, 2013). Erdman (1993) outlined other barriers which likely impinged on the use of self-report measures: (1) it takes time to

administer them in clinic, and (2) they were not a priority under a service delivery model that focused on diagnosis of hearing loss rather than on rehabilitation and the reasons that prompted patients to do something about their hearing. Noble (1978) argues, however that performance tests of hearing loss are insufficient, “off target as regards everyday hearing experience” (p. 236) and not able to measure handicap.

The global move in health services towards patient-centred care, has lent itself to the inclusion of self-report measures (Montano, 2015). As health services have progressed towards a more patient-centred model of care, emphasis on the use of self-report measures in audiology has increased. Since 1964, when the Hearing Handicap Scale (HHS; High, Fairbanks, & Glorig, 1964) was introduced there has been a steady increase in the number and variety of self-report measures available in audiology (Demorest & DeHaven, 1993).

Today, several national audiological societies have implemented the use of self-report measures in routine audiological assessment (American Speech-Language-Hearing Association, 2006; British Society of Audiology, 2012; New Zealand Audiological Society, 2015). There is now more pressure to use self-report measures to demonstrate the efficacy of treatments (Cox, 2005). Currently in New Zealand, a self-report outcome measure is a compulsory part of the documentation required by the Accident Compensation Corporation (ACC) when claiming hearing-aid funding for people with work-related hearing loss (Accident Compensation Corporation, 2014).

### **1.3.3 Self-report measures in audiology.**

In audiology, self-report measures are used to measure perceptions across a variety of domains, including but not limited to hearing disability and handicap

(social and emotional), as well as benefit, performance, satisfaction, and use of amplification devices (Cox, 2005; Erdman, 1993; Tye-Murray, 2015). Disability and handicap related to tinnitus, vestibular disorders, and third party disability can also be measured using self-report measures (Noble, 2013).

Self-report measures may be employed at any stage of audiological assessment or rehabilitation depending on what they are designed to measure. At an initial audiological assessment self-report tools that measure handicap or overall effect of hearing loss on quality of life may be used to supplement audiometric measures and provide a more complete picture of a client's perception of their hearing loss and rehabilitation needs (Weinstein & Ventry, 1983).

Traditionally, self-report measures in audiology have mainly focused on speech perception (Noble, 2013). There are fewer self-report tools that probe perceptions of a broader array of hearing related functions, such as localisation, clarity of non-speech sounds, ability to segregate sounds and identification of distance and direction of movement of sound sources (e.g., Speech Spatial and Qualities of Hearing Scale [SSQ] and the Spatial Hearing Questionnaire [SHQ]; Gatehouse & Noble, 2004; Tyler, Perreau, & Ji, 2009).

#### **1.3.4 Variations of self-report measures.**

There is a sub-group of questionnaires or scales similar to self-report measures, which are completed by someone other than the person with the hearing loss. Some examples include parent- (e.g., Parents Evaluation of Aural/Oral Performance of Children [PEACH]; Ching & Hill, 2007]), teacher- (Teacher's evaluation of aural/oral performance of children [TEACH; Ching, Hill, & Psarros, 2000]), and significant-other report measures (Significant Other Assessment of

Communication [SOAC]; Schow & Nerbonne, 1982). These types of measures can be useful because they may be the only means of assessing the effects of impairment or intervention on an individual (e.g., infants who cannot speak for themselves) or because they provide useful insight (e.g., the report of teachers or significant others).

### **1.3.5 Specific populations.**

Self-report measures are often targeted at specific populations, e.g., infants (LittleEARS; Coninx, Weichbold, & Tsiakpini, 2003), school-aged children (e.g., Listening Inventory for Education [L.I.F.E.]; Anderson & Smaldino, 1996), individuals with cochlear implants (The Nijmegen Cochlear Implant Questionnaire [NICQ]; Hinderink, Krabbe, & Van den Broek, 2000), or the elderly (HHIE; Ventry & Weinstein, 1982). Targeting a specific population can reduce the chance of having irrelevant items in the questionnaire, and in that way produce a more valid measure (Dillon et al., 1997; Weinstein & Ventry, 1983).

### **1.3.6 Open-end versus closed-end questions and response format.**

Open-end and closed-end questions elicit different types of information, and their appropriateness depends on the information required and the intended use of the tool. Open-end questions are useful for eliciting a broad amount of information about a person's circumstances or problems, and are advantageous in that the information is personal and relevant to the individual. They tend to elicit qualitative data, which makes it more difficult to compare responses between individuals in a quantitative manner. While open-ended questions elicit information that is specific and most relevant to a person, there is the potential for responders to get off-topic (Tye-Murray, 2015). Closed-end questions are useful for gathering quantitative data. Closed-end questions can be tailored to allow comparisons between different individuals, and quantify the degree of benefit post an intervention. However, closed-end questions

may restrict people's responses and miss important and relevant information (Tye-Murray, 2015).

Within quantitative data there are two main categories of data that can be collected: categorical and continuous variables. Categorical responses include nominal (e.g., 'yes' or 'no') and ordinal (e.g., ordered categories such as level of education) levels of measurement, and are analysed using frequency and non-parametric statistics (Streiner & Norman, 1995). Continuous variables include interval (interval between values is known) and ratio (has a meaningful zero point; the highest measurement precision) levels of measurement, and can be analysed using parametric statistics (e.g., mean, median, and standard deviation; Froman, 2014; Streiner & Norman, 1995).

The response format of a questionnaire can affect the type of data collected. For example, a questionnaire that gives set responses such as yes, sometimes, or no, (without corresponding numerical values) will give rise to categorical data, while a question that requires the respondents to rate their response on a scale of 0 to 10, elicits a continuous variable. Streiner and Norman (1995) state that it is important to use the most appropriate response format (categorical or continuous) based on the concept being measured. For example, if the question is asking about behaviours or attitudes, it is more appropriate to use a continuum response format than to categorise the responses, because attitudes and behaviours tend to exist as a continuum not as categories with fixed boundaries. Inappropriate response formats can cause greater error in measurement due to reduced response options, and a reduction in correlation to other measures (Streiner & Norman, 1995). Noble (1978) points out that while categorical (or non-scaled) responses can describe the incidence of something in a population, a response format that uses a continuum (i.e., a scaled response) can

describe the incidence as well as the degree to which something is affected, and constitutes “a new source of data” in and of itself that can be compared to data from other tests (p. 244).

### **1.3.7 Limitations of self-report measures.**

Self-report measures for measuring psychological phenomena have been controversial in the past (Jensen & Haynes, 1986), and according to Noble (2013), the debate on the general validity and reliability of self-assessment measures still exists. Some limitations or disadvantages of self-report measures include: poorly framed questions, distortion from contextual factors, time it takes to administer them, and assumptions of equal relevance (Noble, 2013; Dillon et al., 1997). First, poorly framed questions can threaten the validity of a questionnaire, if they do not probe the construct they are supposed to be measuring, or if they are confusing for the person completing the questionnaire (Noble, 2013). Ensuring appropriate readability levels (Jensen & Haynes, 1986), piloting a questionnaire to get feedback from participants on it, and carrying out psychometric measures of reliability, and particularly validity, can help identify questions that are confusing, ambiguous, irrelevant, or not measuring what they were intended to measure.

Second, self-report measures are liable to distortion from contextual factors because answers are up to the discretion of the person filling it out (Noble, 2013). Denial of hearing loss, being unaware of having a hearing loss, financial compensation, and different personalities are just some of the contextual factors that can cause someone to conceal or exaggerate their problems when completing a self-report measure (Cox, 2003; Noble 2013). The method of administration (filling it out by pen-and-paper versus a face-to-face interview) is another contextual factor influencing answers on self-report measures (Noble, 2013). Because self-report

measures are prone to distortion, it is important to know the reliability of a self-report measure across different times and locations, and for different administration methods. Test-retest reliability is a psychometric measure that can assess distortion on a questionnaire.

Third, the time it takes to administer and score self-report measures in the clinic can be a deterrent in their use (Cox, 2003). Screening or short version questionnaires have been developed to reduce the time it takes to administer them and increase the likelihood of them being routinely used in clinic settings (Erdman, 1993; Demeester et al., 2012; Kiessling, Grugel, Meister, & Meis, 2011; Noble, Jensen, Naylor, Bhullar, & Akeroyd, 2013).

Finally, the issue of item relevance can be a disadvantage in self-report measures (Dillon et al., 1997; Gatehouse & Akeroyd, 2007; Noble, 2013). Unless a self-report measure is tailored specifically to each person (e.g., the Client Oriented Scale of Improvement [COSI]; Dillon et al., 1997), some items on a survey may not be relevant to them. Despite this fact, on most inventories with set questions all the items are equally weighted in the scoring, which may reduce the validity of the self-report measure, waste time, and frustrate people who do not feel the scenarios or questions are relevant to their experiences (Dillon et al., 1997). Erdman (1993), points out that patients are more likely to adhere to a recommended treatment if they perceive it to be relevant to what they are experiencing. Self-report measures with fewer items reduces the likelihood of questions being irrelevant, and reduces time and frustration factors for the clinician and client. Regardless of questionnaire length, if it has fixed items the issue of irrelevance cannot be negated. However, in defence of using fixed questionnaires, Noble (2013) argues that

“No one can speak of the experiences of personal disabilities or handicaps better than the person who suffers them. Even if the form of account-making is reduced to numbers on a scaled questionnaire, those numbers are still the result of the person exercising discretion, hopefully with minimum distortion, in choosing a term that most closely matches his or her experience” (Noble, 2013, p. 13).

Given the limitations of self-report measures, certain psychometric measurements of reliability and validity should be assessed in order for a self-report measure to be considered a valid measurement tool.

### **1.3.8 Value of self-report measures.**

Despite the limitations discussed above, the increasingly widespread use of self-report measures across different domains, including audiology, supports their legitimacy and usefulness as an assessment tool (Noble, 2013; Jensen & Haynes, 1986). In advocating the use of self-report measures, Noble (2013) and Erdman (1993) argue in a similar vein two main points: (1) people seek audiological services because they are experiencing perceived difficulties (i.e., disabilities and handicap), therefore their experiences and perceptions related to their hearing ability should be assessed, and (2) despite their limitations, self-report measures are “the only means of assessing the cognitive and affective experiences” of those with hearing loss (Erdman, 1993, p. 307). Jensen and Haynes (1986) add that self-report measures are cost-effective and time efficient, have good face validity, are objectively scored thereby reducing clinician bias or intuitive inference, and can be applied to a wide range of populations and problems. Furthermore, self-report measures are a relatively flexible tool, in that they can be completed at home, or in a waiting room prior to an appointment.



In audiology, self-report measures of disability and/or handicap are recommended, because studies have shown that degree of hearing disability or handicap cannot always be reliably predicted from measures of hearing loss (i.e., pure-tone audiometry) or speech perception ability (i.e., speech audiometry; Erdman, 1993; Gatehouse & Noble, 2004).

Finally, in the real world people are subject to a variety of complex listening environments and situations, which cannot be satisfactorily modelled in a laboratory or clinical environment in order to assess how a person functions in those environments. Therefore, the only way of assessing people's ability to communicate auditorily in complex listening environments is to ask them about it, or have a self-report measure designed to probe those kinds of listening situations (Gatehouse & Noble, 2004).

In summary, self-assessments should be included in audiological assessment for the following reasons:

1. Hearing thresholds do not always reliably predict a person's perceived ability, disability, or handicap.
2. People seek audiological services because they are having trouble hearing in everyday life, therefore it is as appropriate to assess their ability to hear in everyday life using a self-report measure, as it is to measure their hearing thresholds.
3. Self-report measures can probe listening situations that cannot be formally assessed using clinic or laboratory based tests.

## **1.4 Psychometric Issues**

### **1.4.1 Overview.**

Psychometrics refers to “an area of science focusing on the development, use, and interpretation of theoretically sound, consistent and valid assessment instruments...for the systematic measurement of psychological or educational variables”, (McGrath, 2011, p. 1190). There are inherent problems associated with the aim of measuring conceptual ideas such as attitudes and personalities, or hearing disability and handicap. Without considering the psychometric properties of questionnaires, the interpretation of information collected from them is limited. Therefore, an important part of validating a self-report measure is to investigate and comment on its reliability and validity (Johnson & Danhauer, 2002). The following sections will address the concepts of validity and reliability as they relate to self-report measures.

### **1.4.2 Validity.**

In the realm of self-report measures, validity is a measure of the degree to which a questionnaire measures the construct it was designed to measure, i.e., its accuracy (Jensen & Haynes, 1986). For example, if a questionnaire is designed to measure hearing handicap, to know if it is a valid measure one needs to know it accurately measures the concept of hearing handicap and not something else. According to Hyde (2000) there are three preeminent types of validity: construct validity, content validity, and criterion validity. These three types of validity are described below.

#### ***1.4.2.1 Construct validity.***

According to Hyde (2000), a modern way of looking at validity is that there is basically one type of validity, i.e., construct validity, and that the other sub-types of validity contribute to construct validity. Construct validity is an assessment of the degree to which a measure probes the construct it is designed to measure. Construct validity relies on a conceptual framework that outlines the concept or construct a tool is designed to measure, and what variables and relationships are likely to exist between the measure and other measures related to the construct of interest.

Establishing construct validity thereby is a gradual process where studies that explore relationships and patterns between the measure and predicted variables, either confirm or counter expected variations or co-variations, and cumulatively build evidence of its validity or lack thereof (Hyde, 2000). Some tools used for assessing construct validity are factor analysis and correlational analysis (Hyde, 2000). Factor analysis analyses the pattern of responses on items in a scale, and based on their inter-relationships, groups items into factor loaded groups that ideally correspond to the elements or sub-scales differentiated by the designers of the tool. Correlational analysis investigates relationships between the measure and other related measures or predicted variables (Hyde, 2000).

Another common way of assessing construct validity is to conduct large-scale studies comparing groups that differ maximally in the extreme characteristics related to the construct under investigation. The results should support predictions of how the two different groups would perform significantly differently on the scale related to the construct under investigation. For example, a self-report tool designed to measure disability associated with compromised binaural hearing, would be completed by two groups that differ maximally in terms of hearing profile, e.g., those with normal

binaural hearing, and those with monaural hearing or asymmetric hearing loss.

Significant differences in the scores of the two groups would confirm the construct validity of the measure, in that it is sensitive to hearing profile, and behaves as predicted for those with binaural hearing versus monaural hearing.

#### ***1.4.2.2 Content validity.***

Content validity is an assessment of the content of the questions and a critique of their suitability and appropriateness in relation to the constructs they are supposed to measure or probe. Content validity is one of the few psychometric measures that tends to be assessed qualitatively rather than quantitatively, because it relies on the judgments of people, i.e., their opinions of the questions' suitability, and their understanding of the construct they are measuring (Hyde, 2000; Jensen & Haynes, 1986).

Jensen and Haynes (1986) highlight the importance of the method of questionnaire item development and its effect on content validity. They refer to Goldfried and D'Zurilla's (1969) suggested method of item development, whereby situations or problems of interest are made known by getting input from a wide range of representative people, e.g., affected individuals, significant others, and professionals. This ensures a wide range of ideas are considered, the relevant issues are highlighted, and reduces the risk of bias or inclusion of irrelevant items (Jensen & Haynes, 1986).

#### ***1.4.2.3 Criterion validity.***

Criterion validity is an assessment of how the results of a self-report measure compare to results from other assessments thought to measure the same thing (Jensen & Haynes, 1986). For example, results from a self-report tool designed to measure

ability to understand speech in noise, might be compared to scores on a speech in noise test to establish whether there is a relationship where one would expect to see a relationship. Criterion validity is measured by correlating scores from related tests in order to determine covariation (Bentler & Kramer, 2000). Other audiological tests that may be used to assess criterion-related validity include, pure-tone audiometry, speech audiometry, and localisation tests (Bentler & Kramer, 2000). Hyde (2000) cautions against expectations of strong relationships between laboratory measures (e.g., pure-tone audiometry) and disability measures (e.g., self-report measures), stating that the impairment and disability domains have limited overlap.

### **1.4.3 Reliability.**

Reliability refers to how consistent or reproducible a result is (Hyde, 2000). For example, in a game of darts, a reliable player could consistently throw the dart into the same section or quadrant on the board. However, although they may be reliable, their dart throwing may not be accurate (or valid), because they do not hit the bulls-eye. Similarly, a self-report measure can be reliable (consistent) but not valid (accurate), or measuring what it was designed to measure (Johnson & Danhauer, 2002). Noble (2013) states that if a self-report measure is to be used as an outcome measure following intervention, it must demonstrate good reliability in terms of internal consistency and test-retest reliability.

#### ***1.4.3.1 Internal consistency.***

Internal consistency refers to how consistent or correlated scores are between related questionnaire items (Jensen & Haynes, 1986). For example, there may be two questionnaire items that ask questions relating to speech discrimination in noise, but in different situations or contexts. It is expected that these two items would have similar scores (i.e., be correlated or show internal consistency), because if the person

struggles to hear in background noise, and understanding speech in noise is common to both situations, then the answers to both questions are likely to share some variation related to understanding speech in noise. Sub-scales within a questionnaire should demonstrate good internal consistency (Jensen & Haynes, 1986). Cronbach's alpha correlation coefficient is a common measure used to describe internal consistency (the strength of the relationship or amount of shared variance between related items; Jensen & Haynes, 1986). Cronbach's alpha relates the amount of variance in the total scale score to the amount of variance of an individual item. Where there is no shared variance among items, Cronbach's alpha would equal zero, whereas perfectly correlated responses to different questions would yield a Cronbach's alpha of one (Hyde, 2000).

#### ***1.4.3.2 Test-retest reliability.***

##### ***1.4.3.2.1 Overview.***

Test-retest reliability is a measure of how consistent scores are on a questionnaire over time, when no effort is made to affect the dependent variable (Jensen & Haynes, 1986). To assess test-retest reliability, the same group of people complete a particular questionnaire at least twice, with a period of time in between administrations (e.g., 6 weeks) and with no intervention occurring. The scores between the initial completion and the second completion are compared using statistical analyses. With no intervention occurring between the administration times, it is expected that the scores will remain fairly consistent. Test-retest reliability is most often calculated using correlational analysis or a repeated measures ANOVA (analysis of variance) (Streiner & Norman, 1995).

#### *1.4.3.2.2 Importance of test-retest reliability.*

The purpose of assessing test-retest reliability is to quantify the degree to which extraneous factors may influence the results of a questionnaire in the absence of intervention. This is so that when intervention occurs and scores do change, clinicians and patients can tell if the change is clinically significant and attributable to the intervention or not. Test-retest reliability data is also needed to statistically determine what constitutes a significant change in score between administrations (i.e., critical difference scores; Cox, 2003; Noble, 2013). As mentioned earlier in the limitations of self-report measures, a person's answers on a questionnaire can and typically will vary from moment to moment, day to day, week to week, and year to year due to contextual factors. Changes in mental state, situations, circumstances, and ability are some of the variables that can affect self-perceptions at any given moment and therefore affect the reliability of a questionnaire. Test-retest reliability is an important measure, particularly if the self-report is designed to measure benefit over time, so that one can be confident a change is "true" and not just measurement error (Demorest & Walden, 1984).

#### *1.4.3.2.3 How to measure test-retest reliability.*

Test-retest reliability is typically measured by administering the same questionnaire at least twice, with a period of time separating the administrations and no intervention occurring in between. The period of time between the administrations should be representative of the real-life situation in which it will be used.

#### *1.4.3.2.4 Interpreting test-retest reliability results.*

Streiner and Norman (1995) state that reliability is not a property which a self-report measure either has or does not have, but rather it is something that all of them have to a more or less extent. Therefore, the question that remains is "how much

reliability is ‘good enough’?” (Streiner & Norman, 1995, p. 121). Attempts have been made to state what is good enough. Demorest and Walden (1984) state that a value of  $r > 0.80$  and low standard error of measurement is adequate for test-retest reliability. Streiner and Norman (1995) refer to the works of Kelly (1927) and Weiner and Stewart (1984), who recommend a minimum reliability coefficient of 0.94 and 0.85, respectively. However, Streiner and Norman (1995) advise against arbitrary cut-off points and recommend, when deciding on acceptable reliability levels in a study, a consideration of the sample size and population used to calculate the test-retest reliability, and having several replicable studies contribute to the overall test-retest reliability of the measure in question.

Another way of assessing results of test-retest reliability is to compare results against other instrument’s results that are assumed to have acceptable test-retest reliability (Streiner & Norman, 1995). Because there is no literature describing what is acceptable for test-retest reliability in terms of a repeated measures ANOVA, and the fact that most studies of test-retest reliability report correlation coefficients, it is difficult to compare results between correlation and repeated measure ANOVA studies. Demorest and Walden (1984) and Hyde (2000) suggest calculating critical difference scores, because it is more interesting and useful to know the size of the change in score that is required in order for it to be considered a clinically significant change between a pre- and post-intervention administration of a self-report measure.

#### ***1.4.3.3 Factors influencing test-retest reliability.***

Several factors can influence test-retest reliability, some of which can be controlled (e.g., method of administration and time between administrations) and others which are outside the influence of the researcher (e.g., mental state of the person completing the questionnaire). The two main factors influencing test-retest



reliability that will be considered in this study are the time between administrations and method of administration. Descriptions of how these factors can influence test-retest reliability are outlined below.

#### *1.4.3.3.1 Time between administrations.*

Intuitively, one can reason that shorter time intervals between test administrations will increase the reliability of the measure, and this is in fact true (Streiner & Norman, 1995). With a shorter time period between administrations, memory of previous answers could affect responses on following administrations (Hyde, 2000), and there is less opportunity for changed answers as a result of extraneous factors such as changes in mental state, circumstances, ability, and coping strategies. However, it is not going to be an accurate measure of reliability if you assess the test-retest reliability over a 1-week period, when in practice the retest period is likely to be longer (Streiner & Norman, 1995). Conversely, a long time interval between administrations can also adversely affect test-retest reliability measurement in that during a long time period, the attribute may in fact change (e.g., hearing ability may decrease causing an increase in disability). In this case, the change in score may be a “true” change, at which point it is more a measure of stability, rather than test-retest reliability (Demorest & Erdman, 1988; Hyde, 2000).

To ensure measurement accuracy, the test-retest reliability should be assessed using a time-interval as close to the time-interval that the measure is likely to be administered over in real-world practice (Hyde, 2000). Test-retest time periods for some audiology based studies on self-report measures have ranged from 4 to 6 weeks (Fang, Chang, Wan, Wang, & Chen, 2013), 6 weeks (Newman & Weinstein, 1989; Singh & Pichora-Fuller, 2010; Weinstein, Spitzer & Ventry, 1986), 12 weeks (Cox & Rivera, 1992), and up to 40 weeks (Demorest & Erdman, 1988). A time interval of

around 6 weeks seems plausible, as a likely scenario would be 6 weeks between an initial diagnostic assessment, getting hearing aids, and having sufficient time to try the hearing aids. A further 6 weeks beyond that would allow time enough to be adept with using the hearing aids, and gain more experience with the hearing aids in different situations. Cox (2003) reported that some domains of self-report outcome measures appear to stabilise within 3 weeks of the fitting, again suggesting that a period of 6 weeks seems a reasonable time interval to encompass the initial diagnostic assessment through to approximately 3 weeks post-fitting follow-up.

#### *1.4.3.3.2 Administration method.*

Administration method can affect test-retest reliability (Noble, 2013; Streiner & Norman, 1995). Possible administration methods for self-report measures include face-to-face interview, self-administration (pen-and-paper), telephone interview, or computer assisted (Streiner & Norman, 1995). The methods dealt with herein that are currently most pertinent to audiology settings are face-to-face interview and pen-and-paper administration. It has been demonstrated with regards to test-retest reliability, that interview administration tends to yield higher test-retest reliability than pen-and-paper administration (Demorest & Dehaven, 1993; Singh & Pichora-Fuller, 2010; Weinstein, Spitzer, & Ventry, 1986). This is likely due to the fact that assessors can check understanding of questions in a face-to-face interview, and the assessor can elaborate on a question if the person does not understand (Gatehouse & Noble, 2004). In addition, items are less likely to be missed by the respondent. An advantage of pen-and-paper administration is that the client does not have to come into the clinic, freeing up more clinic time for other tasks. However, disadvantages of pen-and-paper administration are that patients may choose not to complete the self-report measure at all, they cannot ask for clarification if they do not understand a question, and they are

more likely to skip questions or not complete the questionnaire correctly (Streiner & Norman, 1995).

#### **1.4.4 Normative data.**

Normative data provides information on average scores on an assessment for a participant sample with particular characteristics. An individual's score becomes more meaningful when it is compared against normative data (Demorest & DeHaven, 1993). Hartman, Roper, and Bradford (1979) identified several purposes for normative data, including: (1) identifying areas that are problematic to a client, (2) establishing realistic rehabilitation goals, (3) separating individuals into groups for treatment or research, (4) comparing subject samples across investigations, and (5) using it as a criterion for evaluating significant clinical changes after treatment. It is useful and important to have comprehensive normative data on self-report measures (Cox, 2003). Normative data can change depending on the characteristics of the sample under investigation, (e.g., people with normal hearing versus experienced hearing-aid wearers), therefore specific sub-group norms should be established (Jensen & Haynes, 1986).

### **1.5 Original SSQ-49 (Speech, Spatial, and Qualities of Hearing Scale)**

#### **1.5.1 Rationale for the SSQ-49.**

Gatehouse and Noble (2004) realised there was a deficit in the complexity of listening situations and types of hearing functions traditionally assessed in hearing disability and handicap self-report questionnaires (Noble, 2010). Questionnaires available at the time tended to focus exclusively on speech understanding, and speech understanding in relatively stationary listening environments. Functions such as localisation and segregation of sounds were not typically assessed, and questions

addressing speech understanding in complex listening environments did not accurately reflect real-life listening experiences where sound sources are constantly varying in location, distance, volume, and pitch (Gatehouse & Noble, 2004). Gatehouse and Noble developed the original Speech, Spatial, and Qualities of Hearing Scale (SSQ-49; Gatehouse & Noble, 2004). They intended for the SSQ to be used for both hearing-aid (acoustic or bone-conduction) and cochlear-implant users, before or after receiving intervention (Gatehouse & Noble, 2004). One of the motivations behind the development of the SSQ-49 was to “fill in the gap” and provide a self-report questionnaire that probed more hearing functions and better addressed complex real-life listening situations. Particular emphasis was given to auditory tasks that implicate binaural function, e.g., localisation, tracking moving sound sources, and understanding speech in dynamic (moving and changing) and noisy environments.

The SSQ-49 consists of 49 questions covering 3 sub-scales. The speech sub-scale has 14 questions, the spatial sub-scale has 17 questions, and the qualities sub-scale has 18 questions. Additional questions are available for the aided condition (Gatehouse & Noble, 2004). The speech sub-scale covers ability to hear speech in a variety of complex listening environments, including in restaurants, background noise, echoic environments, with and without visual cues, and following multiple streams of conversation. The spatial sub-scale focuses on ability to localise speech and non-speech sounds that are stationary or moving, and tell the direction of movement of a travelling sound source. The qualities sub-scale covers a range of characteristics relating to sound and its quality, including naturalness and clarity of sounds, ability to separate out different sound sources occurring simultaneously (i.e., segregation), music identification and clarity, familiar voice recognition, and listening effort (Gatehouse & Noble, 2004).

### **1.5.2 Auditory scene analysis.**

The concepts of auditory scene analysis (Bregman, 1990) and auditory ecology (Gatehouse, Elberling, & Naylor, 1999; Gatehouse, Naylor, & Elberling, 2006b) provide the backdrop against which the SSQ-49 scale was developed (Noble, 2013). Auditory ecology refers to the unique auditory lifestyle or environment of an individual. Noble (2013) uses the example of a professional music teacher and a domestic gardener to illustrate the idea that different people have different auditory environments depending on what activities they engage in on a regular basis. The music teacher may be surrounded by competing noises (music and voices), and engage conversationally with students and others often, while the gardener is in a quiet environment, possibly not talking to many people for much of his day, so their auditory environments and listening demands, or their auditory ecology are quite different.

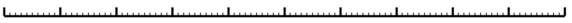
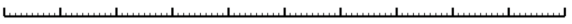
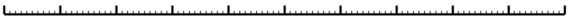
Auditory scene analysis is the process of making sense of one's auditory environment by figuring out which sounds belong to which source (stream segregation) and focusing one's attention on the salient sounds, e.g., speech (Bregman, 1990). In order to make sense of these environments, the auditory system uses cues such as timing or phase differences, loudness differences, and frequency cues to segregate different sound sources using localisation and by perceptually grouping similar sounds together (Bregman, 1990; Moore, 2013; Noble, 2013). The concept of auditory scene analysis highlights two important points: (1) auditory environments are more often than not dynamic, with observers and sound sources constantly moving and altering the sounds arriving at the ear, (2) auditory environments are complex being made up of several different sound sources that are often competing, and require the listener to tease out a signal of interest from a

background of several different competing sounds. Even for a well-functioning auditory system, listening in a complex acoustic environment is challenging, but for someone with an impaired auditory system, it is even more so. Reduced sensitivity and dynamic range, degraded frequency and temporal resolution and asymmetrical hearing all affect one's ability to make sense of sound in a complex auditory environment (Moore, 2013).

Gatehouse and Noble (2004) state that little attention has been given to the complexities of everyday auditory ecology and auditory scene analysis in laboratory-based performance measures and other self-report tools, which have tended to focus on speech recognition in stationary noise. Self-report measures have paid little attention to hearing functions such as segregation and localisation, which rely on normal binaural function (Noble & Gatehouse, 2004). Noble states that the goal of developing the SSQ-49 was to develop a questionnaire that “drills more effectively into the functions of hearing in the everyday world...with emphasis on what the binaural system is called upon to do” (Noble, 2013, p. 106).

### **1.5.3 Description of the SSQ-49.**

The SSQ-49 is a measure of disability associated with hearing loss. The questionnaire uses a scale-response format, whereby every question has a scale rating from 0-10, with 0 indicating much difficulty, and 10 indicating no difficulty. The anchors are always in a negative (left) to positive (right) direction to reduce confusion (Noble, 1998). A ‘not applicable’ option is also included for each question. Patients complete each question by putting a mark somewhere on the scale that represents their perceived ability. Examples of the response format can be seen in Figure 2. Responses are added together both overall and within each sub-scale, to give a total and sub-scale score.

<i>Not at all</i>  0   1   2   3   4   5   6   7   8   9   10	<i>Perfectly</i>	Not applicable <input type="checkbox"/>
<i>Concentrate hard</i>  0   1   2   3   4   5   6   7   8   9   10	<i>Not need to concentrate</i>	Not applicable <input type="checkbox"/>
<i>Lots of effort</i>  0   1   2   3   4   5   6   7   8   9   10	<i>No effort</i>	Not applicable <input type="checkbox"/>

*Figure 2: Examples of response scales in the SSQ-49 (MRC Institute of Hearing Research, n.d.).*

#### **1.5.4 Development of the SSQ-49.**

Gatehouse and Noble mostly developed the SSQ-49 questions themselves but in some instances borrowed or adapted questions from existing questionnaires. The questions in the speech sub-scale were designed to cover a range of realistic contexts of speech hearing, with varying degrees of difficulty, and in varying conditions (Gatehouse & Noble, 2004). Examples include simultaneously listening to someone talking to you while trying to attend to the news on television; following conversation that is rapidly changing between speakers; hearing in echoic versus sound damped environments; following conversation in the presence of speech or environmental noise; and following speech when you can see someone's face versus when you cannot. Situations involving listening on the telephone were also included in the speech section. Some of the questions in the speech section were derived from a questionnaire by Noble (1995), which assessed localisation function and associated handicap, and speech hearing ability.

The spatial section included the assessment of three main components: direction (localisation), distance, and movement of sound in both stationary and dynamic auditory environments (Gatehouse & Noble, 2004). Examples of scenarios

include telling the direction of a speaker when they begin talking; telling the direction of movement of a vehicle, or which direction a person is approaching from by the sound of their footsteps or voice; and one's impression of how far away something is from its sound. The spatial section also includes a question about whether sounds sound like they are in your head or outside of your head - relating to possible occlusion effects from ear moulds.

The final section addresses the qualities of sound, and includes assessment of segregation ability, tone recognition, clarity and naturalness of sounds, and listening effort (Gatehouse & Noble, 2004). Segregation is the ability to discriminate between different sounds (i.e., different sources) that are occurring simultaneously. Recognition refers to things such as identifying a person by the sound of their voice. Clarity and naturalness are assessed with regard to their own voice, others' voices and also the naturalness of everyday non-speech sounds. Listening effort is related to one's ability to follow a conversation and to ignore distracting or competing sounds. Examples of questions include: telling apart a nearby noise from a person speaking to you; clarity and naturalness of other people's voices or music; and concentration or effort needed to follow a conversation. Ability to understand conversation in a car either as the driver or as the passenger is also assessed in the qualities section. Some of the questions addressing recognition and listening effort were derived from the Amsterdam Inventory for Auditory Disability and Handicap (Kramer, Kapteyn, Festen, & Tobi, 1995) and the Profile of Hearing Aid Benefit (PHAB; Cox & Gilmore, 1990), respectively.

The authors piloted the questions on a group of outpatients from the Glasgow Royal Infirmary and made adjustments to the wording of questions where the meaning was ambiguous or confusing (Gatehouse & Noble, 2004). They added new



questions to the qualities section, based on feedback that highlighted relevant topics that were not initially included. The process of piloting and refining the questions was repeated several times to produce the final draft of questions that make up the SSQ-49 (Gatehouse & Noble, 2004).

### **1.5.5 Hearing Handicap Questionnaire (HHQ).**

Gatehouse and Noble (2004) also developed the Hearing Handicap Questionnaire (HHQ), which was designed to be used in conjunction with the SSQ. The HHQ is a 12-item inventory that addresses the personal and social effects of hearing loss (i.e., hearing handicap as opposed to hearing disability covered in the SSQ). It is comprised, in part, of items from two other pre-existing self-report measures: the Glasgow Health Status Inventory section of the Glasgow Benefit Inventory (Robinson, Gatehouse, & Browning, 1996) and the Hearing Disabilities and Handicaps Scale (Hétu, Getty, Philibert, Desilets, & Noble, 1994). All the items included in the HHQ have “demonstrated appropriate leverage in previous applications” according to the authors (Gatehouse & Noble, 2004, p. 88). Each item has a 5-point interval scale with the following response options: never, rarely, sometimes, often, and almost always, with never indicating no associated handicap, and almost always indicating a high degree of handicap. Gatehouse and Noble (2004) demonstrated a correlation of  $r = -0.61$  ( $n = 153$ ) between SSQ and HHQ total average scores for a group of non-hearing-aid wearers. A smaller correlation was found between the HHQ average scores and the better-ear (BEPTA) and worse-ear pure-tone averages (WEPTA;  $r = 0.12$  and  $r = 0.13$ , respectively) for the same sample group, indicating better correlations between hearing disability and handicap than between hearing loss and handicap. Noble et al. (2008) conducted a factor analysis of the HHQ for cochlear implant wearers with three different profiles (unilateral,

bilateral, and bimodal;  $n = 181$ ). They found two main factors: emotional distress (questions 1-7) and social restriction (questions 8-12), which correspond to the two main areas the HHQ was designed to measure, i.e., the personal and social effects of hearing loss.

#### **1.5.6 Primary SSQ-49 studies.**

##### ***1.5.6.1 Gatehouse & Noble, 2004.***

Gatehouse and Noble (2004) describe the rationale and development of the original SSQ-49 and HHQ, and analyse responses on the SSQ-49 for a group of non-hearing-aid wearers ( $n = 153$ , mean age = 71 years,  $SD = 8.1$ ) with a BEPTA of 38.8 dB HL ( $SD = 15.5$ ) and WEPTA of 52.7 dB HL ( $SD = 24.4$ ). They report normative data for the sample (see section 1.7) and data on the relationship between hearing loss, disability and handicap when using the SSQ and HHQ (see section 1.5.5).

##### ***1.5.6.2 Noble & Gatehouse, 2004.***

Noble and Gatehouse (2004) used the same sample as in Gatehouse and Noble (2004), but compared responses on the SSQ-49 between people with symmetrical ( $n = 103$ ) and asymmetrical ( $n = 50$ ;  $\geq 10$  dB HL difference between BEPTA and WEPTA) hearing loss. They found that in general those with asymmetrical hearing loss scored lower on the SSQ, than those with symmetrical hearing loss, particularly in the spatial domain. They conclude that the SSQ is sensitive to departures from normal binaural hearing function.

##### ***1.5.6.3 Noble & Gatehouse, 2006.***

Noble and Gatehouse (2006) compared responses on the SSQ-49 between three groups: pre- hearing-aid fitting ( $n = 144$ ), post-unilateral hearing-aid fitting ( $n = 118$ ), and post-bilateral hearing-aid fitting ( $n = 42$ ). To improve the comparability of

the three groups, they selected those with the best-matched pure-tone thresholds in their better and worse ears resulting in 63 people in the unaided group, 69 in the unilaterally aided group, and 34 in the bilaterally aided group. They found that there were situations where responses showed more benefit with one hearing aid versus no hearing aid (including in the directional component of spatial hearing), and they attributed this to improved audibility. Further situations demonstrated more benefit with two hearing aids versus one hearing aid, particularly in the domains of dynamic spatial hearing (distance and movement), listening effort, and rapidly switching and divided attention.

#### **1.5.7 Research using the SSQ-49.**

Since its development in 2004, the SSQ-49 has been used extensively in a wide variety of research. The majority of the research topics compare binaural and monaural hearing, or bilateral and unilateral aiding with different hearing-aid devices. See Table 1 for a description of the types of research involving the SSQ-49 and references to the studies that have used the SSQ-49.

Table 1: Description of the types of research involving the SSQ-49 and references to the studies that have used the SSQ-49.

Description of research using the SSQ-49	References
Establishing normative data for the SSQ-49	<ul style="list-style-type: none"> <li>• Demeester et al., 2012</li> <li>• Mertens, Punte, &amp; Van de Heyning, 2013</li> <li>• Noble &amp; Gatehouse, 2004</li> </ul>
Establishing reliability or validity of the SSQ-49	<ul style="list-style-type: none"> <li>• Singh and Pichora-Fuller, 2010</li> <li>• Banh, Singh, and Pichora-Fuller, 2012</li> <li>• Moulin and Richard, 2016</li> <li>• Akeroyd, Guy, Harrison, &amp; Suller, 2014</li> <li>• Tyler et al., 2009</li> </ul>
Investigating disability associated with monaural versus binaural hearing, or asymmetric hearing loss	<ul style="list-style-type: none"> <li>• Noble &amp; Gatehouse, 2004</li> <li>• Vannson et al., 2015</li> </ul>
Assessing disability associated with unilateral hearing loss	<ul style="list-style-type: none"> <li>• Douglas, Yeung, Daudia, Gatehouse, &amp; O'Donoghue, 2007</li> <li>• Olsen, Hernvig, &amp; Nielsen, 2012</li> <li>• Vannson et al., 2015</li> </ul>
Comparing outcomes of bilateral versus unilateral hearing aids or cochlear implants	<ul style="list-style-type: none"> <li>• Noble, 2010</li> <li>• Noble &amp; Gatehouse, 2006</li> <li>• Noble, Tyler, Dunn, &amp; Bhullar, 2008b</li> <li>• Noble, Tyler, Dunn, &amp; Bhullar, 2009</li> <li>• Summerfield et al., 2006</li> <li>• Laske, Rösli, Pfiffner, Veraguth, &amp;</li> </ul>

	Huber, 2009
Evaluating the effectiveness of cochlear-implant or hearing-aid features such as noise-reduction algorithms, directional microphones, and frequency compression	<ul style="list-style-type: none"> <li>• Ellis &amp; Munro, 2015</li> <li>• Spriet et al., 2007</li> </ul>
Comparing different modes of hearing, e.g., bone-conduction hearing aids, CROS system hearing aids, and cochlear implants for single-sided deafness	<ul style="list-style-type: none"> <li>• Arndt et al., 2010</li> </ul>
Comparing unilateral, bilateral or bimodal cochlear implant profiles	<ul style="list-style-type: none"> <li>• Dwyer, Firszt, &amp; Reeder, 2014</li> <li>• Farinetti et al., 2015</li> <li>• Potts &amp; Litovsky, 2014</li> </ul>
Using the SSQ-49 as an outcome measure for bilateral and unilateral cochlear implantation	<ul style="list-style-type: none"> <li>• Capretta &amp; Moberly, 2016</li> <li>• Hassepass et al., 2013</li> <li>• Ta Vora-Vieira, Marino, Krishnaswamy, Kuthbutheen, &amp; Rajan, 2013</li> <li>• Zhang et al., 2015</li> </ul>
Using the SSQ-49 as an outcome measure for bone-anchored hearing aids both for bilateral and unilateral hearing loss	<ul style="list-style-type: none"> <li>• McNeil, Gulliver, Morris, Makki, &amp; Bance, 2014</li> <li>• House, Kutz, Chung, &amp; Fisher, 2010</li> <li>• Martin et al., 2010</li> <li>• Pai et al., 2012</li> </ul>
Measuring hearing disability associated with otosclerosis post-surgery	<ul style="list-style-type: none"> <li>• Redfors, Olaison, Karlsson, Hellgren, &amp; Möller, 2015</li> </ul>
Investigating acclimatization effects associated with hearing-aid intervention	<ul style="list-style-type: none"> <li>• Dawes, Munro, Kalluri, &amp; Edwards, 2013</li> </ul>

### **1.5.8 SSQ covariates.**

#### ***1.5.8.1 Age.***

Some studies have shown significant correlations between age and SSQ-49 scores, while others have not. Significant correlations that have been found are in the predicted direction, i.e., negative, indicating that as age increases, scores on the SSQ-49 decrease, which is logical given the increased incidence of hearing loss with older age. Factors that may affect the observation of a significant correlation between age and SSQ-49 scores, include whether the SSQ-49 score is based on aided or unaided condition, the sample size and mean age of the participants involved, and how the participants' SSQ-49 scores and ages are analysed.

Several studies have reported no significant correlations between SSQ-49 scores and age. Capretta and Moberly (2016) found no significant correlation between age and SSQ-49 scores for a group of post-lingually deaf cochlear implant users ( $n = 23$ , mean age = 66.96 years, SD = 9.74). Noble and Gatehouse (2004) found no significant correlation between age and overall SSQ-49 score ( $r \leq -0.03$ ) for a group including people with symmetrical hearing loss ( $n = 103$ , mean age = 70.1 years, SD = 8.3) and asymmetrical hearing loss ( $n = 50$ , mean age = 72.8 years, SD = 7.5). Olsen et al. (2012) found no significant correlation for a group comprised of people with unilateral SNHL ( $n = 98$ , median age = 53 years, range = 18 – 81) and people with hearing within normal limits ( $n = 89$ , median age = 37 years, range = 18 – 66). Likewise, Vannson et al. (2015) found no significant correlations between SSQ-49 scores and age for a group comprised of people with asymmetrical hearing loss ( $n = 49$ , mean age = 47 years, range = 20 – 70) and people with hearing within normal limits ( $n = 11$ , mean age = 40 years, range = 32 – 61).

In comparison, other studies have reported significant correlations between age and SSQ-49 scores. Moulin and Richard (2016) found a significant effect of age for the total SSQ-49 score ( $r = -0.32$ ,  $p < 0.01$ ) and for 9 out of the 10 pragmatic sub-scale scores for a group of people with hearing loss ( $n = 216$ , mean age = 54.2 years,  $SD = 17$ ). Stronger correlations were observed for the speech pragmatic sub-scales ( $-0.31$  –  $-0.37$ ,  $p < 0.01$ ) than the spatial and qualities sub-scales. Studies that compared younger and older age groups using a form of analysis of variance (ANOVA), also found a significant effect of age on SSQ-49 scores. Banh, Singh, and Pichora-Fuller (2012) compared SSQ-49 scores between two groups of people with clinically normal hearing up to 4 kHz; young normal hearing participants (YHN;  $n = 48$ ; age, mean = 18.8 years,  $SD = 1$ ) and older normal hearing subjects (ONH;  $n = 48$ ; age, mean = 70 years,  $SD = 4.1$ ). They conducted a repeated measures analysis of variance (ANOVA) and observed a significant main effect of age ( $p < 0.001$ ), with YNH participants on average scoring higher than ONH participants. Similar results were found by Demeester et al. (2012) who compared 103 participants with normal hearing between 18-25 years of age, with 24 adults aged 55-65 years with clinically normal hearing. They found a significant age-group effect on SSQ-49 scores ( $p < 0.001$ ). The researchers postulated that these results could be due either to small differences in the average pure-tone thresholds of the groups, or threshold-independent age effects, such as auditory processing. Banh et al. (2012) noted that while the adults with normal hearing reported greater disability than their younger counterparts, they still reported on average less hearing disability than people with hearing loss who were of a similar age.

#### ***1.5.8.2 PTA.***

Moderate correlations have been reported between SSQ-49 scores and PTA for the unaided condition. See Table 5 in section 1.6.2.5.

#### ***1.5.8.3 Socioeconomic status.***

Only one study (Capretta & Moberly, 2016) could be found that reported correlations between socioeconomic status and SSQ-49 scores. Capretta and Moberly (2016) measured outcomes on the SSQ-49 for 23 post-lingually deaf CI users. Socioeconomic status was calculated based on occupational status and educational level and ranked from 1 (lowest) to 8 (highest). They found no significant correlation between socioeconomic status ranking and SSQ-49 scores.

#### ***1.5.8.4 Years or level of education.***

Only one study (Moulin and Richard, 2016) could be found that reported correlations between years of education and SSQ-49 scores. Moulin and Richard (2016) found significant correlations between years of formal education and 9 out of the 10 pragmatic scales on the SSQ-49 for a group of 216 participants with hearing loss (mean age = 54.2 years, SD = 17; mean years of formal education = 12.4, SD = 4.3). No significant correlation was found for the listening effort pragmatic sub-scale. The significant correlations ranged from  $r = 0.21$  (localisation) to  $r = 0.32$  (identification and segregation of sound;  $p < 0.01$ ). This indicates that to some degree, higher SSQ-49 scores were linked to more years of formal education. The authors postulate that this result could be related to ability to read and comprehend the questions in the pen-and-paper method of administration.



### **1.5.9 Variations of the original SSQ-49.**

#### ***1.5.9.1 Overview.***

Since the original SSQ-49 was developed in 2004, several variations have been developed to serve particular purposes such as comparing amplification devices, screening for hearing loss, and shorter versions for use in clinical settings. To date, the following variations of the original SSQ-49 are available and will be examined in more detail: SSQ-B, SSQ-C, SSQ-5, SSQ-15, and SSQ-12. The SSQ-49 has also been translated into several different languages (MRC Institute of Hearing Research, n.d.). Most of the SSQ variations mentioned below can be found on the following website: <https://www.ihr.mrc.ac.uk/pages/products/ssq>.

#### ***1.5.9.2 Paediatric versions.***

Paediatric versions of the SSQ were developed by Galvin & Noble (2013) and include a teacher version (SSQ-Teacher), child version (SSQ-Children), and a parent version (SSQ-Parents). The authors suggest administering the questionnaires in an interview format, and recommend 11 years as the minimum age for the child version, and 5 years the minimum age for the teacher and parent versions. The child version consists of 10, 13 and 10 questions for the speech, spatial, and qualities sections, respectively. The parent version consists of 9, 6, and 8 questions for the speech, spatial, and qualities sections, and the teacher version consists of 8, 5, and 8 questions for the respective sections (speech, spatial, and qualities). Modifications to the original SSQ-49 included deletion of irrelevant items, simplifying and changing the wording to make the SSQ more appropriate, and adjustment of scenarios to match children's lifestyles (Galvin, 2013). All the paediatric versions maintain the 10-point scale response format, where 10 equates to perfect ability, and 0 to absolute inability.

#### ***1.5.9.3 Screening version.***

Demeester et al (2012) developed the SSQ-5, which is a screening version of the SSQ-49. It is comprised of five questions from the original SSQ-49, which were selected based on cluster analyses and binary logistic regression analyses. The authors suggest using the short questionnaire to screen for hearing loss, when behavioural screening measures are not available. The screening version consists of 1, 2 and 2 questions from the speech, spatial, and qualities sub-scales, respectively. Studies have demonstrated good correlations between SSQ-5 scores and original SSQ-49 scores (Demeester et al., 2012; Mertens, Punte, and Van de Heyning, 2013; Moulin & Richard, 2016).

#### ***1.5.9.4 Short forms.***

Keissling, Grugel, Mester, and Meis (2011) developed the SSQ-15 (in German) for use in epidemiological settings and as a complementary tool to their laboratory investigations of binaural hearing function. The SSQ-15 includes 5 questions from each sub-scale: speech, spatial, and qualities. Moulin and Richard (2016) found a good correlation between SSQ-49 and SSQ-15 total scores ( $r^2 = 0.97$ ;  $n = 98$  participants with normal hearing,  $n = 216$  participants with hearing loss). Noble et al. (2013) developed the SSQ-12 to provide a more usable version for use in clinical settings where time is limited. The questionnaire is made up of 12 questions selected from the original SSQ-49. A more detailed description of the SSQ-12 is given in section 1.8.

#### ***1.5.9.5 Benefit and comparison versions.***

The SSQ-B (Jensen, Akeroyd, Noble, & Naylor, 2009) is a benefit version tailored for use with first time hearing-aid users. It is designed to measure the benefit afforded by having hearing aids compared to no hearing aids. It is intended that

patients fill out a baseline SSQ (i.e. the original SSQ-49, or SSQ-12) before they receive any intervention, so that the difference between having no hearing aids and then having hearing aids can be measured.

The SSQ-C (Jensen et al., 2009) is a comparison version tailored to compare different hearing aids, or other amplification devices. It is intended that the patient completes a baseline SSQ (SSQ-49 or SSQ-12) for their first hearing aids. The SSQ-C score for the new hearing aids can be compared to the baseline SSQ score to determine if particular hearing aids afford more benefit than others.

The SSQ-B and SSQ-C both use a 10-point scale, but instead of the scale being from 1 to 10, it spans from -5 to +5, where -5 is much worse than without a hearing aid (or with their previous amplification device), 0 is no different, and +5 is much better. The SSQ-B and SSQ-C are available in shorter versions (SSQ-12-B and SSQ-12-C), which use the same questions as in the SSQ-12. As with the longer versions, the SSQ-12-B and SSQ-12-C use the -5 to +5 scale response format.

#### ***1.5.9.6 Different languages.***

The original SSQ-49 as well as the SSQ-B and -C are available in English, Danish, Dutch, German, and Swedish (MRC Institute of Hearing Research, n.d.). Published research using a Korean version of the SSQ-49, indicates that it has also been independently translated into Korean (Heo, Lee, & Lee, 2013; Ryu et al., 2015). Noble in conjunction with the Cochlear corporation also translated the SSQ-49 original into French, Polish, Afrikans, Spanish, Turkish, Italian, and Japanese (MRC Institute of Hearing Research, n.d.). The SSQ-12 as well as the SSQ-12-B and SSQ-12-C are available in English, Danish, and Swedish (MRC Institute of Hearing Research, n.d.).

## **1.6 Psychometric properties of the SSQ-49**

### **1.6.1 Validity of the SSQ-49.**

#### ***1.6.1.1 Factor analysis of the SSQ-49.***

Akeroyd et al. (2014) performed a factor analysis of the SSQ-49 using responses collected from 1220 participants (386 unaided, 627 unilaterally aided, 207 bilaterally aided). They found three clear factors; speech understanding, spatial perception, and clarity, separation, and identification. An emerging fourth factor was effort and concentration. The three clear factors correspond to the three sub-scales that the SSQ-49 is divided into; speech, spatial, and qualities of hearing. The similarities between the three factors identified in the factor analysis and the three sub-scales of the SSQ-49, indicate good construct validity of the SSQ, because it shows good inter-relationship between items that are expected to behave similarly based on their related content (i.e., speech, spatial, or qualities of hearing).

#### ***1.6.1.2 Binaural function.***

Several studies have investigated the effect of unilateral or bilateral aiding, and monaural or binaural hearing on SSQ-49 scores. Given many questions in the SSQ-49 purposefully implicate binaural hearing (Gatehouse & Noble, 2004), it is expected that the SSQ-49 would be sensitive to asymmetrical hearing or differences in unilateral versus bilateral aiding.

Noble and Gatehouse (2004) compared self-ratings on the SSQ-49 between two groups, with symmetrical hearing ( $n = 103$ ) and asymmetrical hearing ( $n = 50$ ). They found the asymmetrical group on average scored consistently lower than the symmetrical group in all three subscales, but more so in the spatial domain. Those with asymmetrical hearing showed more difficulty in the spatial domain on the items

relating to localisation, in the speech domain on items involving following multiple speech streams, and in the qualities domain on items pertaining to segregation of sounds, naturalness, clarity, and conversational effort.

Noble and Gatehouse (2006) compared SSQ-49 ratings for three groups who were unaided ( $n = 63$ ), unilaterally aided ( $n = 69$ ), and bilaterally aided ( $n = 34$ ), with similar hearing thresholds in their better and worse ears. They identified 8 SSQ-49 items where bilateral fittings provided additional benefit compared to a unilateral fitting. Four of those items were from the speech domain, and included complex listening situations such as following multiple speech streams, listening to speech in an echoic environment, and ignoring interfering sounds. Items in the spatial domain included identifying the location or direction of moving sound sources. Bilateral fittings also appeared to provide additional benefit in reducing effort of conversation (quality domain). Situations where bilateral fittings showed significant benefit ( $p < 0.05$ ) and no aid or one aid showed no benefit, were predominantly from the spatial domain, and included locating a dog bark or judging the distance of moving sound sources. In the quality domain bilateral aiding showed benefit for naturalness of their own voice, ability to ignore competing sounds and separate music and voice as separate objects.

Noble, Tyler, Dunn, and Bhullar (2008b) compared SSQ-49 ratings across different CI profiles: one CI ( $n = 70$ ), two CIs ( $n = 36$ ), and a CI with a HA ( $n = 39$ ). A significant effect of implant profile ( $p < 0.05$ ) was found for 6 out of the 10 pragmatic scales (speech in speech contexts, localisation, distance and movement, sound quality and naturalness, identification of sound and objects, and listening effort), and bilateral implantation had the highest ratings across all 10 pragmatic subscales compared to the CI and CI + HA groups. The mean score for the bilateral

implantation group was “strongly distinguishable” (nearly a two scale-point difference) from that of the unilaterally implanted group for the spatial sub-scales. For the speech and qualities sub-scales, the bilateral implant group scored on average approximately one to one-and-a-half points higher than the CI and CI + HA groups. No statistically significant differences were found between the CI and CI + HA profiles.

Laske et al. (2009) compared SSQ-49 ratings for a group of people ( $n = 29$ ) before and after they received a second CI. While they found no statistically significant differences in mean SSQ-49 scores pre and post the second CI, mean and median scores for the bilateral CI condition were consistently higher than with a single CI across all 3 main sub-scales. The spatial sub-scale showed the largest difference in median scores between one versus two CIs, and was the closest to reaching significance ( $p = 0.051$ ).

Overall, results from studies investigating the effect of asymmetrical hearing or unilateral versus bilateral aiding on SSQ-49 profiles suggest that the SSQ-49 is sensitive to deficits in binaural hearing function, or unilateral versus bilateral hearing-aid fittings. Statistically significant effects of symmetrical versus asymmetrical, or unilateral versus bilateral aiding have been found across the three main SSQ-49 sub-scales, but appear to be particularly strong in the spatial domain. Where the results are not statistically significant, researchers at least observed a trend of higher average scores (better ability) for the groups with normal binaural hearing or bilateral hearing-aid fittings. In summary, symmetrical hearing and bilateral hearing-aid fittings “offer advantage[s] for challenging conditions (e.g., multistream signal monitoring), for dynamic spatial hearing, and in reducing listening effort” (Noble, 2010, p. 568), as well as better segregation ability, as seen in the studies reviewed above.

### ***1.6.1.3 Localisation tests.***

This section reports on three studies involving CI wearers that have investigated the relationship between objective localisation tests and self-reported ability on the SSQ-49. All three studies found strong correlations between objective tests of localisation ability (minimum audible angles) and self-reported ability on the SSQ-49 (see Table 2). Zhang et al. (2015) found significant correlations between the objective localisation test and SSQ-49 total scores and spatial sub-scale scores, while Noble et al. (2008b) and Laske et al (2009) only found significant correlations between the localisation test and the spatial sub-scale scores (see Table 2). Noble et al. (2008b) noted that while the correlations they found were low, there was a clear and logical pattern linking localisation performance to self-ratings in the spatial sub-scales, but not to the speech and qualities pragmatic sub-scales. The same trend was seen in the results from Laske et al. (2009). Noble et al. (2008b) concluded that the weak link between self-reported ability on the SSQ-49 and laboratory performance tests of speech recognition and localisation is likely due to the limitations of laboratory tests to recreate the complexities of hearing environments and situations encountered in real life, thus rationalising the use of self-report measures. The large differences in correlations reported by Noble et al. (2008b) compared to Zhang et al. (2015), when both studies used the Everyday Sounds Localisation test, may be attributed to differences in CI profile. Zhang et al. (2015) only included bilateral CI recipients, whereas the study by Noble et al. (2008b) included unilateral and bimodal CI profiles.

Table 2: Studies showing a relationship between the SSQ-49 and localisation tests (significant correlations are in bold).

	Localisation test	n	Hearing aid device	Correlation measures	r value
<b>Laske et al., 2009.</b>	12 speakers in circular array, signal presented at 65 dB SPL.	29	CI + CI	SSQ spatial x LT	<b>Sig. corr.</b> (p < 0.05).
<b>Zhang et al., 2015</b>	Everyday Sounds Localisation test	19 pre-op, 11 post-op	CI + CI	SSQ total x LT	r = 0.051 (pre-op) – <b>- 0.870**</b> (post-op)
				SSQ spatial x LT	r = -0.104 (pre-op) – <b>- 0.854**</b> (post-op)
<b>Noble et al., 2008b.</b>	Everyday Sounds Localisation test	49	CI, CI + CI, CI + HA	SSQ localisation pragmatic sub-scale x LT	r = <b>- 0.35*</b>
				SSQ distance and movement pragmatic sub-scale x LT	r = <b>- 0.31*</b>
				SSQ speech & quality pragmatic sub-scale x LT	No sig. corr.

*Note.* \* significant at p < 0.05 level. \*\* significant at p < 0.01 level. LT = localisation test. Sig. corr. = significant correlation. CI = cochlear implant. HA = hearing aid. Everyday Sounds Localisation Test – Signal presented at 70 dB C, 8 loudspeakers spaced 15.5° apart forming a 108° arc. Pre-op = before cochlear implantation. Post-op = after cochlear implantation. BEPTA and WEPTA not reported for any of the studies.



#### ***1.6.1.4 Other self-reports.***

Four studies investigated the relationship between scores on the SSQ-49 and other hearing-related self-report measures in the disability and handicap domains. The questionnaires used included the HHQ (Gatehouse & Noble, 2004), the HHIE/A (Weinstein & Ventry, 1982; Newman et al., 1990), the Spatial Hearing Questionnaire (SHQ; Tyler et al., 2009), and the Nijmegen Cochlear Implant Questionnaire (NCIQ; Hinderink et al., 2000). With the exception of the asymmetric group in Noble and Gatehouse (2004), all the different self-report measures showed statistically significant strong correlations with the SSQ-49 ( $r \geq 0.71$ ; see Table 3). Particularly strong correlations ( $r = 0.79 - 0.935$ ) were found between total and sub-scale scores on the SSQ-49 and SHQ (Spatial Hearing Questionnaire; the only other disability specific measure in Table 3) indicating good construct validity (Tyler et al., 2009; Zhang et al., 2015). Capretta and Moberly (2016) found a weaker correlation between the HHIE/A total score and the SSQ-49 spatial sub-scale score ( $r = -0.51$ ,  $p < 0.01$ ), suggesting that spatial hearing ability has less impact on hearing handicap than speech perception and qualities of hearing.

Table 3: Studies showing correlations between the SSQ-49 and other hearing-related self-report measures (handicap and disability).

	n	Hearing aid device	BEPTA (dB HL) (SD)	WEPTA (dB HL) (SD)	Self-report measure correlations	r value
Noble & Gatehouse, 2004	103	None	38.9 (15.9)	42.3 (15.7)	SSQ x HHQ total scores	r = <b>-0.77**</b>
	50 (AHL)	None	38.7 (14.9)	74.2 (25.2)	SSQ x HHQ total scores	r = <b>-0.28**</b>
Tyler et al., 2009	139	CI, CI + CI	-	-	SSQ x SHQ total scores.	r = <b>0.79**</b>
					SSQ x SHQ spatial scores	r = <b>-0.84**</b>
Zhang et al., 2015	19 (11 at 6 months)	CI + CI	-	-	SSQ x SHQ total scores.	r = <b>0.790**</b> (24 months post-op) – <b>-0.935**</b> (6 months post-op)
					SSQ x SHQ speech scores	r = <b>0.714**</b> (pre-op) – <b>0.929**</b> (6 months post-op)
					SSQ x SHQ spatial scores	r = <b>0.790**</b> (24 months post-op) – <b>0.908**</b> (12 months post-op)
Capretta & Moberly, 2016	23	CI, CI + CI, or CI + HA	89.8 (post-op)	-	SSQ x HHIE/A total scores	r = <b>-0.71**</b>
					SSQ x NCIQ total scores	r = <b>-0.72**</b>

*Note.* \*\* significant at  $p < 0.01$  level. AHL = asymmetric hearing loss. CI = cochlear implant. HA = hearing aid. BEPTA = better ear pure-tone average .5 – 4 kHz (dB HL). WEPTA = worse ear pure-tone average .5 – 4 kHz (dB HL). SD = standard deviation. SSQ = Speech, Spatial, and Qualities of Hearing Scale. HHQ = Hearing Handicap Questionnaire. SHQ = Spatial Hearing Questionnaire. HHIE/A = Hearing Handicap Inventory for the Elderly/Adults. NCIQ = Nijmegen Cochlear Implant Questionnaire (health-related quality of life measure encompassing hearing, speech, psychological, and social domains).

#### ***1.6.1.5 Speech perception tests.***

Four studies were identified that investigated correlations between speech perception tests and self-rated ability on the SSQ-49. Weak to strong correlations were found between objective speech perception tests and self-reported ability on the SSQ (see Table 4).

Noble et al. (2008b) compared performance on a word recognition test with self-rated ability on the SSQ-49 for a group of CI recipients. Performance on the word recognition test was significantly correlated ( $p < 0.05$ ) with all 10 of the pragmatic sub-scales ( $r = 0.31 - 0.46$ ; see Table 4). The general trend showed lower correlations for the spatial sub-scales than for the speech and qualities sub-scales. Noble et al. (2008b) attributed the weak correlations between the speech sub-scales and the word recognition test to the poor face validity of the word recognition test. The test does not include visual input and is therefore not likely to be a reliable predictor of how people with severe to profound hearing loss cope in real life.

Capretta and Moberly (2016) found significant correlations between the SSQ-49 speech sub-scale and speech perception tests (see Table 4). No significant correlations were found between the speech perception tests and the SSQ-49 total score, and spatial and qualities sub-scale scores. The weakest correlation was with the spatial sub-scale.

Vannson et al. (2015) investigated the relationship between a speech-in-noise test and the SSQ-49 for participants ( $n = 49$ ) with asymmetric hearing loss who did not wear hearing aids. They found statistically significant correlations in the dichotic condition (signal at poorer ear, noise at better ear) for the SSQ-49 total and sub-scale scores, with the strongest correlation for the speech sub-scale (see Table 4). No significant correlations were found for the diotic (signal and noise presented in front)

and reversed dichotic conditions (signal at better ear, noise at poorer ear), which may be due to the fact that they found the dichotic condition to be the most sensitive measure for establishing speech recognition thresholds in noise.

Zhang et al. (2015) found statistically significant correlations for post-cochlear implantation measures only. The correlations at 6-months post-cochlear implantation are larger than the correlations found in the other studies. Again this could be related to the fact that participants in their study all had bilateral CIs, whereas the other studies included participants with unilateral and bimodal CIs (Capretta and Moberly, 2016; Noble et al., 2008b) and asymmetrical hearing (Vannson et al., 2015).

Overall, the studies show some degree of significant correlations between speech perception tests and SSQ-49 scores. While the correlations with SSQ-49 speech scores are not necessarily clearly distinguishable from the total or qualities scores, the spatial sub-scale consistently showed lower correlations. The fact that SSQ-49 speech sub-scale ratings do not show strong correlations with objective speech perception tests, speaks to the limitations of laboratory based tests (reduced face validity), and conversely the value of self-report measures for measuring real-world ability in a range of dynamic speech listening contexts.

Table 4: Studies reporting on correlations between SSQ-49 scores and speech perception tests.

	n	Hearing aid device	BEPTA (dB HL)	WEPTA (dB HL)	Speech test	Correlated measures	r value
<b>Noble et al., 2008b</b>	64	CI, CI + CI, CI + HA	-	-	CNC in quiet.	SSQ 10 pragmatic scales x WRQ	r = <b>0.31*</b> (speech in quiet) – <b>0.46***</b> (multiple speech stream processing and switching)
<b>Zhang et al., 2015</b>	11-19	CI + CI	-	-	CNC in quiet.	SSQ total x WRQ	r = 0.188 (PreCI) – <b>0.839**</b> (PstCI)
						SSQ speech x WRQ	r = 0.238 (PreCI) – <b>0.772**</b> (PstCI)
<b>Vannson et al., 2015</b>	49 (AHL)	None	14 (SD = 8.59)	57.72 (SD = 20.06)	FMT in noise - dichotic condition.	SSQ total x SRN	r = <b>-0.38</b> , p = 0.01
						SSQ speech x SRN	r = <b>-0.40</b> , p = 0.006
						SSQ spatial x SRN	r = <b>-0.32</b> , p = 0.047
						SSQ qualities x SRN	r = <b>0.32</b> , p = 0.047
<b>Capretta &amp; Moberly, 2016</b>	23	CI, CI + CI, CI + HA	88.2 (pre-op)	-	CID-22 in quiet.	SSQ speech x WRQ	r = <b>0.56*</b>
	16		89.8 (post-op)		AzBIO in quiet.	SSQ speech x SRQ	r = <b>0.61*</b>

Note. \*significant at p < 0.05 level. \*\* significant at p < 0.01 level. \*\*\* significant at p < 0.001 level. NS = not significant. AHL = asymmetric hearing loss. CI = cochlear implant. HA = hearing aid. SD = standard deviation. FMT = French Matrix Test in noise; dichotic listening condition (signal at poorer ear, noise at better ear). SRN = sentence recognition in noise. SRQ = sentence recognition in quiet. WRQ = word recognition in quiet. 10 PS – SSQ's 10 pragmatic scales. PreCI = Pre-cochlear implant. PstCI = 6 months post-cochlear implant. CID-22 = The Central Institute for the Deaf-22 word recognition test. AzBIO = sentence recognition in quiet test. CNC = consonant-nucleus-consonant words. SSQ = Speech, Spatial, and Qualities of Hearing Scale.

#### ***1.6.1.6 Hearing ability.***

Correlations between self-report scores and hearing thresholds is seen as evidence of construct validity for self-report measures, as it is expected there will be some relationship between self-reported hearing disability and degree of hearing loss (Demorest & DeHaven, 1993). With regards to the SSQ-49, some studies have reported significant correlations between PTA (.5-4 kHz) and SSQ-49 scores of up to  $r = -0.56$  (see Table 5). A negative correlation is consistent with comparing PTA and SSQ-49 scores given a larger PTA (poorer hearing) is likely to result in lower scores on the SSQ-49 (greater disability).

Table 5 shows results from studies that have investigated the relationship between hearing thresholds and SSQ-49 scores. The results are variable, with some studies (Gatehouse & Noble, 2004; Moulin & Richard, 2016; Noble & Gatehouse) reporting significant moderate correlations, and others reporting no significant correlations between hearing thresholds and SSQ-49 scores (Olsen et al. 2012; Singh & Pichora-Fuller, 2010). Singh & Pichora-Fuller (2010) suggest their failure to find a significant correlation may be related to differences in sample group characteristics, because their sample groups had significantly better mean hearing thresholds and a smaller range of thresholds, compared to Gatehouse and Noble (2004), Noble and Gatehouse (2004), and Noble and Gatehouse (2006) (see Table 5). Banh et al. (2012) also failed to find any significant correlation between PTA and SSQ-49 scores for two groups (younger adults,  $n = 48$  and older adults,  $n = 48$ ) with essentially normal hearing up to 3 kHz in both ears, except for a weak correlation between the spatial sub-scale score and the PTA of the older group ( $r = 0.32$ ,  $p < 0.05$ ).

Noble and Gatehouse (2006) correlated scores on individual SSQ-49 items with BEPTA for 3 threshold-matched groups, with either no HA, one HA or two HAs.

They found more significant correlations for the group with moderate hearing loss but no HAs, than in the group with one HA, and found no significant correlations for the group with two HAs. Noble and Gatehouse (2006) argue that this is a logical pattern of results given hearing aids are supposed to reduce the disability associated with hearing loss, thereby increasing scores on the SSQ-49 and weakening the link between PTA and self-reported hearing disability.

Moulin and Richard (2016) investigated the relationship between PTA and SSQ-49 scores for 216 participants with hearing loss. Moulin and Richard (2016) noted that mean SSQ-49 scores decreased as hearing loss increased (a similar pattern was noted by Akeroyd et al. 2014). Moulin and Richard (2016) found similar sized moderate correlations when compared to Gatehouse and Noble (2004), Noble and Gatehouse (2004; symmetrical hearing group only) and Noble and Gatehouse (2006; unaided group only), and did not show marked differences between BEPTA and WEPTA correlations. Noble and Gatehouse (2006), however, did find a difference between BEPTA and WEPTA, for the asymmetric hearing loss group, where BEPTA x SSQ-49 was significantly correlated, but WEPTA x SSQ-49 was not. Moulin and Richard (2016) performed a step-wise regression analysis to find the variables that best predicted SSQ-49 scores. They found BEPTA ( $\beta = -0.63$ ) and hearing loss asymmetry ( $\beta = -0.25$ ) were the best predictors of SSQ-49 scores.

Olsen et al. (2012) investigated self-reported disability for people with asymmetric hearing loss, but failed to find any significant relationship between SSQ-49 scores and PTA, however they did not specify what PTA they were correlating the scores with (i.e., BEPTA, WEPTA, or PTA of both ears) so their results are inconclusive. Contrary to Noble and Gatehouse (2004), Vannson et al. (2015) found a significant correlation for SSQ-49 x WEPTA, but not for BEPTA (see Table 5). This

may be the case due to differences in BEPTA, where Noble and Gatehouse's (2004) participants had a moderate hearing loss in their better ear, while participants in Vannson et al. (2015) study had normal hearing in their better ear.

Capretta and Moberly (2016) assessed a group of CI recipients and found a significant correlation with the spatial sub-scale and total score for the post-implant condition only, however they only measured the SSQ-49 post-implant, so that is a possible reason why the post-implant SSQ-49 scores did not correlate with the pre-implant BEPTA.

Overall, the pattern of results for asymmetric hearing loss appears less conclusive, but moderate symmetrical hearing loss in the unaided condition appears to show a reliable correlation to SSQ-49 scores (Gatehouse & Noble, 2004; Noble & Gatehouse, 2004; Noble & Gatehouse, 2006), and normal hearing or mild hearing loss appears to show no significant correlations to SSQ-49 scores (Singh & Pichora-Fuller, 2010; Olsen et al., 2012). There is evidence of PTA being related to SSQ-49 scores, providing some degree of validity to the self-report measure. Given the rationale for using self-report measures is because thresholds do not sufficiently predict disability and handicap, it is reasonable to find only weak to moderate correlations between disability and handicap and PTA.



Table 5: Studies reporting on the relationship between SSQ-49 scores and hearing ability (PTA).

	n	Hearing aid device	BEPTA (dB HL) (SD)	WEPTA (dB HL) (SD)	Correlation measures	<i>r</i> or $\beta$ value
<b>Gatehouse &amp; Noble, 2004</b>	153	None	38.8 (15.5)	52.7 (24.4)	SSQ total x BEPTA	<i>r</i> = <b>-0.51*</b>
					SSQ total x WEPTA	<i>r</i> = <b>-0.52*</b>
<b>Noble &amp; Gatehouse, 2004</b>	103	None	38.9 (15.9)	42.3 (15.7)	SSQ total x BEPTA	<i>r</i> = <b>-0.43**</b>
					SSQ total x WEPTA	<i>r</i> = <b>-0.40**</b>
	50 (AHL)	None	38.7 (14.9)	74.2 (25.2)	SSQ total x BEPTA	<i>r</i> = <b>-0.55**</b>
					SSQ total x WEPTA	<i>r</i> = -0.21
<b>Noble &amp; Gatehouse, 2006</b>	63	None	52.7 (6.7)	66.7 (17.9)	SSQ speech items x BEPTA	<i>r</i> = -0.07 – <b>-0.41**</b>
					SSQ spatial items x BEPTA	<i>r</i> = 0.02 – <b>-0.33**</b>
					SSQ qualities items x BEPTA	<i>r</i> = -0.05 – <b>-0.47**</b>
	69	HA	56 (7.3)	68.5 (10.5)	SSQ speech items x BEPTA	<i>r</i> = -0.01 – -0.21
					SSQ spatial items x BEPTA	<i>r</i> = 0.00 – <b>-0.27*</b>
					SSQ qualities items x BEPTA	<i>r</i> = 0.00 – <b>-0.35**</b>
	34	HA + HA	54.9 (13.5)	66.5 (16.6)	SSQ speech items x BEPTA	<i>r</i> = 0.01 – -0.21
					SSQ spatial items x BEPTA	<i>r</i> = 0.00 – -0.32
					SSQ qualities items x BEPTA	<i>r</i> = 0.02 – -0.26
<b>Singh &amp; Pichora-Fuller, 2010</b>	139	None	14.4 (7.4) – 20.9 (9.4)†	21.2 (10.6) – 26.8 (10.9)†	SSQ Total x BEPTA	<i>r</i> < ± 0.20
<b>Olsen et al., 2012</b>	98 (AHL)	15% wore HA/s	< 20 at .25-1 kHz, <30 at 2 kHz, and < 40 at 4 kHz	> 80 at .25-4 kHz	SSQ speech x PTA	<i>r</i> = -0.039
					SSQ spatial x PTA	<i>r</i> = 0.079
					SSQ qualities x PTA	<i>r</i> = -0.048
	89	None	< 20 at .25-1 kHz, <30 at 2 kHz, and < 40 at 4 kHz		SSQ speech x PTA	<i>r</i> = -0.113
					SSQ spatial x PTA	<i>r</i> = 0.084
					SSQ qualities x PTA	<i>r</i> = 0.057

Vannson et al., 2015	49 (AHL)	None	14 (8.59)	57.72 (20.06)	SSQ total x BEPTA	r = -0.06, p 0.6227 F(1, 44) = 0.135, p > 0.05
					SSQ total x WEPTA	r = <b>-0.34**</b> , p = 0.0079 F(1, 44) = 4.99, p < 0.05
Capretta & Moberly, 2016	19	CI, CI + CI, CI + HA	88.2 (pre-op)	-	SSQ Total x BEPTA	β = 0.31
					SSQ speech x BEPTA	β = 0.19
					SSQ spatial x BEPTA	β = 0.33
					SSQ qualities x BEPTA	β = 0.25
	23		89.8 (post-op)	-	SSQ Total x BEPTA	β = <b>0.48*</b>
					SSQ speech x BEPTA	β = 0.38
					SSQ spatial x BEPTA	β = <b>0.47*</b>
					SSQ qualities x BEPTA	β = 0.38
Moulin & Richard, 2016	216	9.3% wore HA/s	26 (15) with 15.2 (SD = 23.3) dB HL of asymmetry	-	SSQ Total x BEPTA	r = <b>-0.56**</b>
					SSQ Total x WEPTA	r = <b>-0.52**</b>
					SSQ speech x BEPTA	r = <b>-0.57**</b>
					SSQ speech x WEPTA	r = <b>-0.43**</b>
					SSQ spatial x BEPTA	r = <b>-0.47**</b>
					SSQ spatial x WEPTA	r = <b>-0.56**</b>
					SSQ qualities x BEPTA	r = <b>-0.49**</b>
					SSQ qualities x WEPTA	r = <b>-0.44**</b>

*Note.* \* significant at  $p < 0.05$  level. \*\* significant at  $p < 0.01$  level. SD = standard deviation. PTA = pure-tone average of 0.5 – 4 kHz (for the Olsen study it was not clear from which ear the PTA for the correlation was taken). BEPTA = better ear pure-tone average, 0.5 – 4 kHz unless otherwise stated. WEPTA = worse ear pure-tone average, 0.5 – 4 kHz unless otherwise stated. AHL = asymmetric hearing loss. HA = hearing aid. CI = cochlear implant. † = lower and upper ranges of PTA from the 4 groups included in the study.  $\beta$  = linear regression analysis. SSQ = Speech, Spatial, and Qualities of Hearing Scale.

### **1.6.2 Reliability of the SSQ-49.**

To date, reliability measures of the SSQ-49 have been recorded in two studies. Singh and Pichora-Fuller (2010) calculated the test-retest reliability of the SSQ-49 and compared interview and pen-and-paper administration methods with older people with no experience with hearing aids, and calculated internal consistency. Demeester et al. (2012) also calculated internal consistency of the SSQ-49 with groups of younger and older, and participants with hearing loss and normal hearing.

#### ***1.6.2.1 Test-retest reliability.***

Singh and Pichora-Fuller (2010) assessed the test-retest reliability of the original SSQ-49 (Noble & Gatehouse, 2004) for four sample groups using different combinations of administration methods: interview and pen-and-paper (mail). The four groups were interview-interview, interview-mail, mail-interview, and mail-mail. They administered the SSQ-49 twice (with a 6-month time interval) to 159 participants, including 97 females and 62 males, aged 60 to 88 years (mean = 72.8; SD = 5.6), who had essentially normal hearing and no experience with hearing aids. Pearson-product moment correlation scores were calculated for the total and sub-scales scores between the two administration times for each administration group. They found the total and sub-scale scores between the first and second administrations were stable across the four administration groups, with all total and sub-scale scores showing significant and high correlations ( $r = 0.49 - 0.86$ ,  $p < 0.01$ ). They found the highest total score correlation for the interview-interview group ( $r = 0.83$ ,  $p < 0.01$ ), and the lowest total score correlation for the mail-mail group ( $r = 0.65$ ,  $p < 0.01$ ), which is consistent with previous findings that have demonstrated higher test-retest reliability for the interview administration method than the pen-and-

paper administration method (Demorest & Dehaven, 1993; Weinstein et al., 1986).

The largest between-administration group score change for total and sub-scale scores was 0.4. The mean between-administration participant score change on the total SSQ-49 score was 0.7 (SD = 0.6). The full test-retest correlation results for the mail-mail group (n = 40, PTA mean = 21.8 dB HL, SD = 11) were as follows;  $r = 0.65$  (total score),  $r = 0.83$  (speech),  $r = 0.56$  (spatial), and  $r = 0.64$  (qualities), which were all significant at the  $p < 0.01$  level. Apart from the interview-interview group, the general trend showed lower test-retest correlations for the spatial sub-scale (see Table 6).

A potential drawback with the methodology of Singh and Pichora-Fuller's (2010) test-retest reliability study is the length of time between administrations (6 months). The longer period between administrations may have caused poorer test-retest reliability than would be found when the SSQ-49 is administered within a shorter time period similar to how it would typically be used in real-world practice (e.g., 6 weeks between pre- and post-intervention administrations).

Table 6: Test-retest correlations of the SSQ-49 (total and sub-scale scores) for four different combinations of administration methods (n = 39 for the interview – mail group, and n = 40 for the remaining groups). Adapted from Singh and Pichora-Fuller (2010).

Administration method	Total	Speech	Spatial	Qualities
Interview – Interview	0.83**	0.77**	0.86**	0.83**
Interview – Mail	0.66**	0.71**	0.49**	0.65**
Mail – Interview	0.69**	0.74**	0.59**	0.69**
Mail – Mail	0.65**	0.83**	0.56**	0.64**

Note. \*\* correlation significant at  $p < 0.01$  level (2 tailed).

Overall, the SSQ-49 test-retest correlations reported by Singh and Pichora-Fuller (2010) are generally consistent with test-retest reliability data from other audiology based self-report tools (see Table 7). The lower test-retest correlations for the spatial sub-scale may be due to the fact that participants were unable to clarify the

meaning of questions with an interviewer, or because they may not have paid close attention to their ability to localise a dog bark, or judge the distance of an object by its sound, given they had normal hearing.

Table 7: Test-retest reliability data from hearing-related self-report measures for individuals with hearing loss (except some in the SSQ study who had normal hearing) using the pen-and-paper administration method.

Self-report measure and data source	n	Aided/ Unaided	Time between administrations	No. of administrations	No. of items	Test-retest reliability data
<b>HHIE</b> – Weinstein et al., 1986.	27	Not stated	6 weeks	2	25	$r = 0.84^*$ , $p < 0.01$
<b>HHIA</b> – Newman et al., 1991.	28	Unaided	6 weeks	2	25	$r = 0.97^*$
<b>HHIA-S</b> – Newman et al., 1991.	28	Unaided	6 weeks	2	10	$r = 0.93^*$
<b>APHAB</b> – Cox & Alexander, 1995.	27	Unaided	12 weeks	3	24	( $F[2, 52] = 3.04$ , $p = 0.06$ )
<b>APHAB</b> – Cox & Alexander, 1995.	27	Unaided	12 weeks	2	24	$r = 0.65$ (RVS) – $0.89$ (AVS)
<b>APHAB</b> – Cox & Alexander, 1995.	27	Aided	12 weeks	2	24	$r = 0.67$ (BNS) – $0.81$ (RVS)
<b>Chinese version APHAB</b> – Kam, Tong, & van Hasselt, 2011.	27	Aided	2-4 weeks	2	24	$r = 0.84$ (ICC) and $0.73^{**}$
<b>PHAP</b> – Cox & Gilmore, 1990.	30	Aided	10-20 days	2	66	$r = 0.66$ (RCS) – $0.88$ (BNS)
<b>PHAB</b> – Cox & Rivera, 1992.	28	Aided	12 weeks	3	66	( $F[2,54] = 0.56$ , $p > .05$ )
<b>PHAB</b> – Cox & Rivera, 1992.	33	Aided	12 weeks	3	66	$r = 0.42$ (DSS) – $0.72$ (AVS) ***
<b>Chinese version SADL</b> – Fang et al., 2013.	39	Aided	4-6 weeks	2	15	$r = 0.98^{**}$ , $p \leq 0.05$
<b>SADL</b> – Cox & Alexander, 1999.	72	Aided	12-30 weeks	2	14	$r = 0.81$
<b>Short version HAPI</b> – Schum, 1993.	64	Aided	34 days	2	38	$r = 0.80^*$ , $p < .001$
<b>Danish IOI-HA</b> – Jespersen, Bille, & Legarth, 2014.	25	Aided	1 year	2	7	$0.75^{**}$ , $p < 0.01$
<b>SSQ</b> – Singh & Pichora-Fuller, 2010.	40	Unaided	6 months	2	46	$r = 0.65$ (total), $0.83$ (speech), $0.56$ (spatial), $0.64$ (qualities), $p < 0.01$

Note. \* Pearson product moment correlation. \*\* Spearman's correlation coefficients. \*\*\* Mean  $r$  for both the 3 and 6 month administration times.

ICC = Intra-class correlation coefficient. RVS = Reverberation sub-scale. AVS = Aversiveness sub-scale. BNS = Background noise sub-scale. RCS = Reduced cues sub-scale. DSS = Distortion of sounds sub-scale. HHIA-S = HHIA screening version. APHAB = Abbreviated profile of hearing aid benefit. PHAP = Profile of hearing aid performance. PHAB = Profile of hearing aid benefit. SADL = Satisfaction with amplification in daily life. HAPI = Hearing aid performance inventory. IOI-HA = International outcome inventory for hearing aids. SSQ = Speech, Spatial, and Qualities of Hearing Scale.

#### ***1.6.2.2 Internal consistency.***

Singh and Pichora-Fuller (2010) assessed the internal consistency of the original SSQ-49. They calculated Cronbach's alpha for the overall and sub-scale scores for each administration group at both administration times. They found the SSQ-49 had high internal consistency ( $\geq 0.88$ ) across the four different administration groups at both administration times, based on a generally agreed upon cut-off value for internal consistency of 0.7 (Nunnally, 1978). Demeester et al. (2012) confirmed the high internal consistency results obtained in Singh and Pichora-Fuller's study (2010). They found Cronbach's alpha scores of  $\geq 92\%$  across the total and sub-scale SSQ-49 scores, using the self-administration method ( $n = 127$  clinically normal hearing,  $n = 109$  hearing impaired).

Mertens et al. (2013) assessed the internal consistency of the SSQ-49 for adult CI recipients ( $n = 53$  unilateral,  $1 =$  bilateral). Similar to Singh and Pichora-Fuller (2010), they found a high degree of internal consistency (Cronbach's alpha = 0.94, 0.94, and 0.92 for the speech, spatial, and qualities sub-scales, respectively).

### **1.7 SSQ-49 Normative data**

The following normative data has been reported from studies involving people with and without hearing loss. The normative data provides realistic targets for rehabilitation for different hearing profiles. Note that even young people with normal hearing do not rate their abilities at 100% (see Table 8). Normative SSQ-49 data for young people with hearing within normal limits is reported in Banh et al. (2012), and Demeester et al. (2012), and both studies report similar values (see Table 8). Zahorik and Rothpletz (2015) calculated normative data for young people with hearing within normal limits, but it is not included in Table 8 below, because the data were presented graphically rather than numerically. Normative SSQ-49 data for older people with

hearing within normal limits is reported by Singh and Pichora-Fuller (2010; see Table 9), and Banh et al. (2012) and Demeester et al. (2012; see Table 8). Demeester et al. (2012) calculated the following cut-off scores for the SSQ-49, below which scores indicate significant disability: 6.84 (speech), 6.14 (spatial), 8.18 (quality), and 7.25 (total). The cut-off scores were derived from the mean scores of 103 young people with normal hearing aged 18-25 years, and were taken at 2 SD below the group mean scores on the SSQ-49.

Gatehouse and Noble (2004), Moulin and Richard (2016), and Demeester et al. (2012) report normative SSQ-49 data for older people with hearing loss. Participants in Gatehouse and Noble (2004) were unaided and < 21 out of the 216 participants in Moulin and Richard (2016) wore a hearing aid in at least one ear. It is not stated in the Demeester et al. (2012) study, whether the people with hearing loss wore hearing aids or not. A comparison of SSQ-49 scores between older and young people with hearing within normal limits (Banh et al., 2012; Demeester et al., 2012) shows consistently lower scores for the older group (see Table 8). There is more variation in the normative data reported between the studies involving participants with hearing loss, than participants with hearing within normal limits (see Table 8). This is likely due to greater variations in PTA between the different samples of people with hearing loss. Noble, Naylor, Bhullar, and Akeroyd (2012) report four categories of normative data for older adults with hearing loss: clinic baseline (unaided), stratified sample baseline (unaided and report hearing difficulty and poor self-assessed ability), hearing-aid outcome (bilaterally aided), and ideal outcome (no perceived hearing difficulty and high self-assessed ability ratings). Results from that study have not been included in Table 8 below, because they were presented graphically rather than numerically, and data were reported for the ten pragmatic sub-



scale scores rather than for the total and three main sub-scale scores. Noble and Gatehouse (2006) also report SSQ-49 scores for groups of unaided, unilaterally aided, and bilaterally aided, but the mean scores are given for each item, rather than the mean total, and three main sub-scale scores.

Normative data for CI recipients pre- and post-implantation is reported by Zhang et al. (2015) for bilateral CI recipients, and by Távora-Vieira et al. (2015) and Mertens et al. (2013) for unilateral CI recipients (post-implantation scores only). Vannson et al. (2015) calculated normative SSQ-49 data for people with asymmetric hearing loss and compared it to cut-off points outlined by Demeester et al. (2012). The data were graphically represented using box plots, so are not included in Table 8. House et al. (2010) reported normative SSQ-49 data in graphical form for people with unilateral hearing loss who either have a bone-conduction hearing aid or no bone-conduction hearing aid. The results from the formerly mentioned studies including the normative data for CI recipients and asymmetrical hearing loss are not included in Table 8, but are mentioned here for the purpose of indicating what data are available and where it may be obtained.

Table 8: Normative SSQ-49 data from Banh et al., 2012; Gatehouse & Noble, 2004; Demeester et al., 2012; and Moulin & Richard, 2016.

	Gatehouse & Noble, 2004		Moulin & Richard, 2016		Demeester et al., 2012						Banh et al., 2012			
	(n = 153) HI		(n = 216) HI		(n = 109) HI		(n = 103) YNH		(n = 24) ONH		(n = 48) ONH		(n = 48) YNH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SSQ total score	5.6	2.6	6.6	2.0	7.7	2.1	8.8	0.79	8.1	8.8	7.7	1.2	8.8	0.6
SSQ speech score	4.4	2.4	5.9	2.3	7.2	2.5	8.7	0.91	7.8	-	7.3	1.5	8.5	0.7
SSQ spatial score	5.6	2.6	6.5*	-	7.6	2.0	8.5	1.17	7.9	-	7.5	1.6	8.6	0.7
SSQ qualities score	6.8	2.7	7.2	1.9	8.2	1.8	9.3	0.57	8.7	-	8.3	1.1	9.4	0.6
WEPTA (dB HL)	52.7	24.4	-	-	F.I. < 20 (n=37), 20 ≤ F.I. ≤ 30 (n=30), F.I. > 30 (n=42)	≤ 25 .125-8 kHz				PTA <sub>.25-3 kHz</sub> ≤ 25				
BEPTA (dB HL)	38.8	15.5	26	15										
Asymmetry (dB HL)	-	-	15.2	23.3	-	-	-	-	-	-	-	-	-	-
Age (years)	71	8.1	54.2	17.0	62.9	5.3	19.4	1.3	62.7	2.6	70	4.1	18.8	1.0

*Notes.* BEPTA = better ear pure-tone average of 0.5 – 4 kHz unless otherwise stated (dB HL). WEPTA = worst ear pure-tone average of 0.5 – 4 kHz unless otherwise stated (dB HL). Asymmetry in dB HL. \* approximately. F.I = Fletcher Index (dB HL) calculated from left ear audiogram (PTA<sub>.5-2 kHz</sub>). ONH = older normal-hearing adults. YNH = younger normal-hearing adults. HI = hearing impaired. SSQ = Speech, Spatial, and Qualities of Hearing Scale.

Table 9: Normative SSQ-49 data from Singh & Pichora-Fuller, 2010.

	Singh & Pichora-Fuller, 2010							
	(n = 40) I-I		(n = 39) I-M		(n = 40) M-I		(n = 40) M-M	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<b>SSQ</b> total score	7.5	0.8	8	1.1	7.1	1.4	7.6	1.2
<b>SSQ</b> speech score	7	1.3	7.5	1.6	6.4	1.9	6.9	1.8
<b>SSQ</b> spatial score	7.2	1.5	8	1.4	7	1.8	7.7	1.4
<b>SSQ</b> qualities score	7.4	0.8	8	1.1	7.1	1.4	7.6	1.2
<b>WEPTA</b> (dB HL)	21.2	10.6	22.3	13.4	26.8	10.9	24.7	12.6
<b>BEPTA</b> (dB HL)	17.8	8.5	14.4	7.4	20.9	9.4	18.8	11.3
<b>Age (years)</b>	72.4	5.3	71.6	5.6	73.2	5.9	73.8	5.5
<b>Education (years)</b>	15	3.4	15.9	3.9	14.3	2.9	14.1	3.1

*Notes.* I-I = interview-interview group, I-M = interview-mail group, M-I = mail-interview group, M-M = mail-mail group. SD = standard deviation. SSQ = Speech, Spatial, and Qualities of Hearing Scale. WEPTA = worse ear pure-tone average of 0.5 – 4 kHz, BEPTA = better ear pure-tone average of 0.5 – 4 kHz.

## 1.8 SSQ-12

### 1.8.1 Introduction.

The SSQ-12 is a short form of the SSQ-49, developed for use in clinical or research settings, where time is constrained. The developers' aim in making the SSQ-12 was to “compile a set of items that represent the scale as a whole, offering the clinician or researcher an efficient scaled-down version” (Noble et al., 2013, p. 2).

### 1.8.2 Description of the SSQ-12.

The SSQ-12 is comprised of 12 items selected from the original SSQ-49. It uses the same response format as the original SSQ-49, i.e., a scale from 0 to 10, where 10 equals perfect ability, and 0 equals no ability. A ‘not applicable’ option is given for each item. Questions 1-5 are from the speech sub-scale, 6-8 from the spatial, and 9-12

from the qualities sub-scale. A full copy of the SSQ-12 can be found on the following website: <https://www.ihr.mrc.ac.uk/pages/products/ssq>.

### **1.8.3 Development of the SSQ-12.**

Noble et al. developed the SSQ-12 in 2013. Three independent centres (Eriksholm, MRC Institute of Hearing Research, University of New England) were involved in selecting the SSQ-12 items through a process of deliberation. First, each centre independently chose 12 items they thought best represented the SSQ-49 as a whole, then the three centres compared and discussed their chosen items to come up with the final 12 items included in the SSQ-12 (Noble et al., 2013).

The inclusion of 12 questions in the SSQ-12 was influenced by two factors: (1) the number of items in the SSQ-12 matched the number of items in the Hearing Handicap Questionnaire (HHQ; Gatehouse & Noble, 2004) which was developed to be used in conjunction with the SSQ, and (2) the number of items was close to the number of pragmatic scales (10) determined by Gatehouse and Akeroyd (2006; Noble et al., 2013). Nine out of the ten pragmatic scales are included in the SSQ-12, excluding the speech in quiet pragmatic scale.

### **1.8.4 Research using the SSQ-12.**

Apart from the studies comparing the SSQ-12 and SSQ-49 scores (Mertens & Van de Heyning, 2016; Moulin & Richard, 2016; and Noble et al., 2013) only one published study could be found that has used the SSQ-12 as an outcome measure (Eberhand, Olsen, Miyazaki, Bille, & Caye-Thomasen, 2016). Eberhand et al. (2016) used the SSQ-12 as a subjective outcome measure for 12 patients unilaterally implanted with a new transcutaneous bone-conduction hearing aid. The SSQ-12 was completed once at 6 months post-implantation. They found higher sub-scale scores for

the qualities section, and lower scores on the spatial section and a trend of higher SSQ-12 scores for the people with CHL (n = 5) and MHL (n = 3), than those with single-sided deafness (SSD; n = 4).

### **1.8.5 Psychometric properties of the SSQ-12.**

#### ***1.8.5.1 Validity.***

##### *1.8.5.2.1 Comparison between SSQ-12 and SSQ-49 scores, and other short versions.*

Three studies have compared SSQ-12 scores to SSQ-49 scores: Mertens and Van de Heyning (2016), Moulin and Richard (2016), and Noble et al. (2013). Mertens and Van de Heyning compared scores between the short and long SSQ versions for a group of CI recipients (n = 54) using Spearman's correlations and found a good agreement between the SSQ-12 and SSQ-49 scores ( $r^2 = 0.90$ ). They found a poorer agreement between the SSQ-5 and SSQ-49 scores ( $r^2 = 0.62$ ), concluding that the SSQ-12 is a more valid measure than the SSQ-5 for use in clinical settings where a short version of the SSQ is most appropriate.

Noble et al. (2013) compared SSQ-12 and SSQ-49 scores for a large sample (n = 1220) which included unaided (n = 386), unilaterally aided (n = 627), and bilaterally aided (n = 207) participants. They used a power function to compare scores between the SSQ-12 and SSQ-49, and found "a close agreement" between the SSQ-12 and SSQ-49 scores (p. 4). They observed "modestly lower" average scores on the SSQ-12 compared to the SSQ-49, and a slightly steeper slope on the power function linking the two versions scores, indicating "greater contrast between lower and higher self-ratings on the short version compared with the complete version" (p. 4).

Moulin and Richard (2016) compared SSQ-12 scores with SSQ-49 scores for a sample including participants with normal hearing ( $n = 98$ ) and hearing loss ( $n = 216$ ) using linear regression and power functions. They reported a high correlation between the SSQ-12 and SSQ-49 scores ( $r^2 = 0.96$ ), which is similar to results reported by Mertens and Van de Heyning (2016;  $r^2 = 0.90$ ). Moulin and Richard also found high correlations between the SSQ-49 and SSQ-5 ( $r^2 = 0.90$ ), and SSQ-49 and SSQ-15 ( $r^2 = 0.97$ ). Similar to Noble et al. (2013) they reported significantly lower SSQ-12 scores compared with SSQ-49 scores, but found no significant differences between the SSQ-49, SSQ-5, and SSQ-15 scores. For ease of comparison with Noble et al.'s (2013) results, they reported power function equations linking the SSQ-12 and SSQ-49 scores, which are similar to those reported by Noble et al. (2013). Overall, the data comparing SSQ-12 and SSQ-49 scores suggests that the SSQ-12 is a valid alternative tool for assessing hearing disability, when use of the longer SSQ-49 version is not feasible.

#### ***1.8.5.2 Reliability.***

##### *1.8.5.1.1 Test-retest reliability.*

To my knowledge, as of yet there have been no published studies on the test-retest reliability of the SSQ-12.

##### *1.8.5.1.2 Internal consistency.*

To my knowledge, as of yet there have been no published studies on the internal consistency of the SSQ-12.

#### **1.8.6 SSQ-12 Normative Data.**

To my knowledge, as of yet there have been no published studies reporting normative data for the SSQ-12.

## **1.9 Gap in the literature**

To date there is limited data available on the SSQ-12 to add to its validity, and no studies to date have published normative data or reported on its reliability. The purpose of this current study is to investigate the test-retest reliability of the SSQ-12 for hearing-aid wearers using a pen-and-paper administration method.

## **1.10 Research question**

Is the SSQ-12 a reliable questionnaire for measuring hearing-aid outcomes for hearing-aid wearers when using the pen-and-paper administration method across 3 administration times: T0, T1, and T2?

## **1.11 Hypotheses**

There will be no significant differences between the SSQ-12 scores for each participant's T0, T1, and T2 data for the speech sub-scale.

There will be no significant differences between the SSQ-12 scores for each participant's T0, T1, and T2 data for the spatial sub-scale.

There will be no significant differences between the SSQ-12 scores for each participant's T0, T1, and T2 data for the qualities sub-scale.

There will be no significant differences between the SSQ-12 total scores for each participant's T0, T1, and T2 data.

## Chapter Two: Methods

### 2.1 *A priori* power analysis

For a repeated measures ANOVA design an *a priori* sample size of 27 was calculated using G\*Power 3.1 software with the following parameters: effect size  $f = 0.06$ ,  $\alpha = 0.05$ , power = 0.80, correlation among repeated measures = 0.5, and non-sphericity correction = 1. For a correlation, an *a priori* sample size of 27 was calculated with the following parameters: one-tailed test, effect size = 0.447 ( $r^2 = 0.2$ ),  $\alpha = 0.05$ , and power = 0.80.

### 2.2 Ethics approval

Ethics approval for this study was obtained on 5<sup>th</sup> August 2015 from the University of Canterbury Human Ethics Committee (see Appendix A).

### 2.3 Participants

The study included 28 participants, 17 male and 11 female, with a mean age of 63.89 years (SD = 15.86, range 22 - 83).

#### 2.3.2 Recruitment.

Participants were recruited from the University of Canterbury Speech and Hearing Clinic database, and from the general public. Those recruited from the Speech and Hearing Clinic database had opted to be contacted for research studies and were contacted initially by mail (with the information and consent forms) and then by a phone call a week later. Those recruited from the general public were contacted by poster advertisements (see Appendix B for a copy of the poster) put up in various public spaces, and by word-of-mouth.



### **2.3.3 Inclusion/exclusion criteria.**

To be eligible for inclusion in the study, participants had to have a hearing loss (PTA .5 – 4 kHz > 20 dB HL) in at least one ear, currently use a hearing aid or hearing aids for a minimum of 6 months, and be aged 18 years or older. There were no exclusion criteria, so that the results could be more generalisable.

### **2.3.4 Drop-outs.**

Originally 29 participants were recruited and began the study. One participant dropped-out after T0 for undisclosed reasons. Their data has been excluded from all statistical analyses, including the descriptive statistics of the sample.

## **2.4 Measures**

### **2.4.1 Pure-tone audiometry.**

Air conduction pure-tone audiometry was carried out for frequencies between 250 - 8000 Hz using the threshold-seeking procedures outlined in the University of Canterbury Audiology Protocols and Guidelines or the New Zealand Audiological Society Best Practice Guidelines (New Zealand Audiological Society, 2016; University of Canterbury Speech and Hearing Clinic, 2015). Participants who had a recent audiogram (carried out  $\leq 6$  months prior to the study) were not retested, but instead sent their latest audiogram to the researcher for data collection. Better ear pure-tone average (BEPTA) and worse ear pure-tone average (WEPTA) were calculated for octave frequencies between 500 - 4000 Hz. Pure-tone audiometry was carried out using a calibrated Grason-Stadler GSI-61 audiometer, and Telephonics TDH-39 supra-aural earphones or Etymotic ER-3A insert earphones. Testing was completed in a sound-treated booth with ambient noise levels in compliance with ANSI S3.1 (1999) standards.

### **2.4.2 Demographic questionnaire.**

The demographic questionnaire was made up specifically for this study, and included questions on age, gender, ethnicity, relationship status, household income, and highest level of education. The final question asked participants to rate the severity of their hearing problem on a scale of 1 to 10 (1 = not at all severe, 10 = very severe). See Appendix D for a copy of the demographic questionnaire.

### **2.4.3 Hearing Handicap Questionnaire (HHQ).**

As previously mentioned in section 1.5.5, the Hearing Handicap Questionnaire (HHQ; Gatehouse & Noble, 2004) is comprised of 12 questions probing the personal and social effects of hearing loss. The HHQ was designed for use in conjunction with the SSQ-49 and SSQ-12. The HHQ was sent out with the information and consent forms and completed once at the beginning of the study either at home or at the University of Canterbury Speech and Hearing Clinic. In either case, the HHQ was self-administered. The responses ('never', 'rarely', 'sometimes', 'often', and 'almost always') were assigned 1, 2, 3, 4, and 5 points respectively for analysis purposes. An example of the types of questions in the HHQ is "How often does your difficulty with your hearing affect the way you feel about yourself?"

### **2.4.4 Speech, Spatial and Qualities of Hearing Scale – short version (SSQ-12).**

For a full description of the SSQ-12 see section 1.8. The SSQ-12 was printed on 4 sides of paper, the first page had the instructions on how to complete the questionnaire and space for details about the date of completion, participant age, if participants wore hearing aids, and if so which ear(s), and how long they had worn the hearing aid(s). The twelve questions were on the remaining 3 pages, with four

questions per page. As mentioned in section 1.8.2, the SSQ-12 response format is a visual analogue scale ranging from 0 to 10, where 0 equals no ability and 10 equals perfect ability. The scale is marked down to 0.5 point spaces. Participants are instructed to put a mark such as a cross (X) anywhere on the scale that represents their self-rated ability. If marks were placed between the 0.5 point spaces, scores were estimated to the nearest 0.25 points (e.g. 1.75 or 5.25). To check the reliability of the scoring, half of the T0, T1, and T2 data were randomly selected and scored by an independent rater. Scores were compared between the two raters using Cronbach's alpha and intra-class correlation coefficients. Results showed the scoring of the SSQ-12 was reliable – Cronbach's alpha ranged from 0.993 – 1.0, and intraclass correlation coefficients (ICC) ranged from 0.987 – 1.0, indicating the two independent raters had excellent agreement beyond chance (Fliess, 1981).

## **2.5 Procedure**

Once participants gave consent to take part in the study (see Appendix C for a copy of the information and consent forms), they completed the demographic questionnaire and HHQ, and pure-tone audiometry was carried out for individuals without a recent audiogram. The demographic questionnaire and HHQ were either completed at home or at the University of Canterbury Speech and Hearing Clinic when they came in for their hearing test. Once the demographic, HHQ and pure-tone data were collected for each individual, they were sent by mail a copy of the SSQ-12 (T0), with a personal note asking them to fill it out on or close to a particular day, and return it to the researcher in the provided return envelope. The SSQ-12 was sent out a total of three times (T0, T1 and T2) to each participant. The time between administrations was 6 weeks (42 days), and was calculated from the date of completion marked on their previous SSQ-12 form. See Table 10 for the mean time

between administrations. Note the number of SSQ-12 responses used to calculate the mean time between administrations varies. This is because one SSQ-12 response from the T0 data, and three SSQ-12 responses from the T2 data did not have the date of completion recorded on the form. The SSQ-12 responses with no date of completion on the form were excluded from the analysis of mean time between administrations where appropriate.

Table 10: Mean, standard deviation, and range of days between administration times.

Administration times	N	Days between administration times		
		<i>Mean</i>	<i>SD</i>	<i>Range</i>
<b>T0 – T1</b>	27	43.47	2.96	41 – 55
<b>T1 – T2</b>	25	44.52	4.5	40 – 61
<b>T0 – T2</b>	24	88	7.32	82 – 116

## 2.6 Statistical Analyses

Statistical analyses were computed using IBM SPSS Statistics version 23 for Macintosh computers. Possible covariates were determined using Pearson's correlation coefficients and Chi Square. Test-retest reliability of the SSQ-12 was determined using a repeated-measures analysis of variance (ANOVA), with time as the within subjects factor, and controlling for any covariates that occurred. The 95% confidence intervals and critical difference scores for the SSQ-12 were calculated according to the methods described by Demorest and Walden (1984).

## Chapter Three: Results

### 3.1 Overview

There were no significant differences between the SSQ-12 total scores for each participant's T0, T1, and T2 data, as well as for the speech, spatial, and qualities sub-scale scores, indicating the SSQ-12 has good test-retest reliability for hearing-aid wearers using the pen-and-paper administration method.

### 3.2 Participant description

The study included 28 participants, 17 male and 11 female, with a mean age of 63.89 years (SD = 15.86, range 22 - 83). The sample had a mean BEPTA of 39.69 dB HL (SD = 16.06, range 11.25 - 80), and a mean WEPTA of 48.08 dB HL (SD = 16.14, range = 23.75 - 83.75). At the commencement of the study, 24 out of the 29 participants (82.76%) had a bilateral hearing-aid fitting and 5 (17.24%) had a unilateral hearing-aid fitting. See Table 11 below for the mean HHQ scores, self-rated severity of hearing loss scores, and years of hearing-aid use. See Figure 3 for the mean pure-tone audiogram for the left and right ears of the sample.

Table 11: Mean, standard deviation and range values for participants' age, self-rated severity of hearing loss, HHQ scores, BEPTA, WEPTA, and years of HA use (N = 28).

	Mean	SD	Range
Age (years)	63.86	15.86	22 – 83
Severity rating (1-10)	5.3	1.8	1 – 8.5
HHQ mean score	2.43	0.65	1.58 – 4.17
BEPTA (dB HL)	39.69	16.06	11.25 – 80
WEPTA (dB HL)	48.08	16.15	23.75 – 83.75
HA use (years)	7.87	9.86	0.5 – 48

*Note.* HHQ = Hearing Handicap Questionnaire. BEPTA = better ear pure-tone average of 0.5 – 4 kHz. WEPTA = worse ear pure-tone average of 0.5 – 4 kHz. HA = hearing aid.

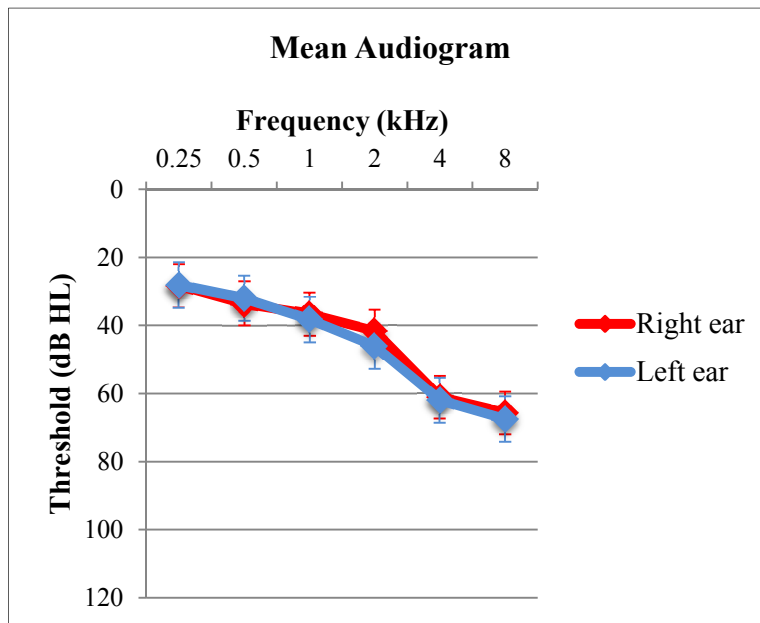


Figure 3: Mean audiogram of participants for right and left ears, with standard error of measurement indicated by the vertical bars (N = 28).

The majority of participants (85.71%) were New Zealand European. The sample percentages of ethnicity, income bracket and level of education are shown in Table 12. Table 13 shows the minimum, maximum, mean, and standard deviation values for the samples total and sub-scale SSQ-12 scores for T0, T1 and T2 data.

Table 12: Sample percentages of ethnicity, income, and level of education (N = 28).

	Percentage of sample
<b>Ethnicity</b>	
NZE	85.71 %
Maori	7.14 %
Other	7.14 %
<b>Income</b>	
< \$50K	50 %
> \$50K	50 %
<b>Level of education</b>	
< HS	21.43 %
HS	10.71 %
Certificate or diploma	17.86 %
BA, post-graduate diploma, MA, PhD	50 %

Note. NZE = New Zealand European. HS = high school. BA = bachelor degree. MA = masters. PhD = doctorate.

Table 13: Minimum, maximum, mean and standard deviation of the SSQ-12 scores for total and sub-scale scores from T0, T1, and T2 data (N = 28).

	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>SD</b>
<b>T0 Total</b>	1.77	9.58	5.667	1.756
<b>T1 Total</b>	2.48	9.36	6.027	1.762
<b>T2 Total</b>	1.67	8.94	5.752	1.803
<b>T0 Speech</b>	1.85	9.10	5.259	1.642
<b>T1 Speech</b>	2.05	9.00	5.500	1.998
<b>T2 Speech</b>	1.60	8.25	5.430	1.667
<b>T0 Spatial</b>	0.33	10.00	5.721	2.339
<b>T1 Spatial</b>	1.67	10.00	6.243	2.162
<b>T2 Spatial</b>	0.33	9.67	5.649	2.359
<b>T0 Qualities</b>	0.00	9.88	5.944	2.483
<b>T1 Qualities</b>	2.63	9.38	6.506	1.967
<b>T2 Qualities</b>	1.25	9.50	6.246	2.054

*Note.* T0 = administration one. T1 = administration two. T2 = administration three. SD = standard deviation.

Significant correlations ( $p < 0.01$ ) of interest were found between WEPTA and self-rated severity of hearing loss, mean HHQ scores and self-rated severity of hearing loss, and between mean HHQ scores and WEPTA (see Table 14).

Table 14: Correlation matrix for self-rated severity of hearing loss, BEPTA, WEPTA, and mean HHQ scores (N = 28).

	<b>Severity</b>	<b>BEPTA</b>	<b>WEPTA</b>	<b>HHQavg</b>
<b>Age</b>	0.032	0.254	0.078	-0.230
<b>Severity</b>	-	0.356	<b>0.573**</b>	<b>0.512**</b>
<b>BEPTA</b>	-	-	<b>0.808**</b>	0.088
<b>WEPTA</b>	-	-	-	<b>0.512**</b>

*Note.* \*\* correlation is significant at  $p < 0.01$  level (2-tailed). HHQ = Hearing Handicap Questionnaire. BEPTA = better ear pure-tone average of 0.5 – 4 kHz. WEPTA = worse ear pure-tone average of 0.5 – 4 kHz.

### 3.3 Correlations between SSQ-12 scores and participant characteristics

(covariates)

Pearson correlations between participant characteristics (continuous variables only) and SSQ-12 scores for T0, T1, and T2 data showed significant correlations ( $p < 0.05$ ) only for mean HHQ score, and self-rated severity of hearing loss (see Table 15).

Table 15: Pearson correlation coefficients of mean SSQ-12 scores, and possible covariates (age, self-rated severity of hearing loss, BEPTA, WEPTA, and mean HHQ scores). Significant correlations are indicated in bold.

	n	Age	Severity	BEPTA	WEPTA	HHQavg
<b>T0 Total</b>	27	0.134	-0.251	0.068	-0.114	<b>-0.386*</b>
<b>T1 Total</b>	28	0.144	-0.231	0.173	-0.039	<b>-0.427*</b>
<b>T2 Total</b>	28	0.12	-0.298	0.095	-0.1	-0.36
<b>T0 Speech</b>	28	0.28	-0.064	0.066	-0.088	-0.341
<b>T1 Speech</b>	28	0.166	-0.203	0.024	-0.122	<b>-0.421*</b>
<b>T2 Speech</b>	28	0.085	-0.277	-0.059	-0.219	-0.344
<b>T0 Spatial</b>	28	-0.069	<b>-0.455*</b>	-0.052	-0.192	-0.31
<b>T1 Spatial</b>	28	0.001	-0.191	0.294	0.141	-0.206
<b>T2 Spatial</b>	28	0.107	<b>-0.380*</b>	0.141	-0.043	-0.318
<b>T0 Qualities</b>	27	0.13	-0.167	0.118	-0.066	-0.371
<b>T1 Qualities</b>	28	0.165	-0.208	0.191	-0.063	<b>-0.439*</b>
<b>T2 Qualities</b>	28	0.145	-0.167	0.189	-0.001	-0.318

Note. \* correlation is significant at  $p < 0.05$  level (2-tailed). HHQavg = Hearing Handicap Questionnaire average. BEPTA = better ear pure-tone average of 0.5 – 4 kHz. WEPTA = worse ear pure-tone average of 0.5 – 4 kHz. T0 = administration one. T1 = administration two. T2 = administration three.

### 3.4 Test-retest reliability

Results of the repeated measures ANOVA controlling for self-rated severity of hearing loss and mean HHQ score, showed no significant changes over time for the SSQ-12 total score,  $F(2,48) = 0.300$ ,  $p = 0.742$ ,  $\eta^2 = 0.012$ . There were also no significant changes over time for the speech,  $F(2,50) = 0.811$ ,  $p = 0.450$ ,  $\eta^2 = 0.031$ , spatial,  $F(2,50) = 0.677$ ,  $p = 0.513$ ,  $\eta^2 = 0.026$ , and quality,  $F(2,48) = 0.597$ ,  $p = 0.554$ ,  $\eta^2 = 0.024$ , sub-scale scores. The observed power for the repeated measures



ANOVA results was 0.095, 0.181, 0.158, and 0.144 for the total, speech, spatial, and qualities sections, respectively. The results of the repeated measures ANOVAs indicate good test-retest reliability of the SSQ-12 for this sample and administration method. Table 16 shows Pearson correlations between the mean total and sub-scale scores for T0, T1, and T2 data. Test-retest reliability correlation values are highlighted in red bold.

Table 16: Pearson's correlation matrix of T0, T1, and T2 data for total SSQ-12 scores, and speech, spatial, and qualities sub-scale scores. The correlations in red bold indicate the test-retest reliability of the SSQ-12 using Pearson correlation values for the total, speech, spatial, and qualities sections across administration times T0-T1, T0-T2, and T1-T2.

	T0Speech	T0Spatial	T0Qualities	T1Total	T1Speech	T1Spatial	T1Qualities	T2Total	T2Speech	T2Spatial	T2Qualities
<b>T0Total (N)</b>	0.871** 27	0.815** 27	0.951** 27	<b>0.791**</b> 27	0.640** 27	0.705** 27	0.737** 27	<b>0.765**</b> 27	0.618** 27	0.713** 27	0.790** 27
<b>T0Speech (N)</b>	1 28	0.454* 28	0.782** 27	0.601** 28	<b>0.571**</b> 28	0.411* 28	0.541** 28	0.506** 28	<b>0.444*</b> 28	0.39* 28	0.552** 28
<b>T0Spatial (N)</b>	- 28	1 28	0.705** 27	0.633** 28	0.384* 28	<b>0.793**</b> 28	0.558** 28	0.733** 28	0.520** 28	<b>0.813**</b> 28	0.699** 28
<b>T0Qualities (N)</b>	- 27	- 27	1 27	0.805** 27	0.647** 27	0.666** 27	<b>0.797**</b> 27	0.762** 27	0.606** 27	0.687** 27	<b>0.816**</b> 27
<b>T1Total (N)</b>	- 28	- 28	- 28	1 28	0.889** 28	0.708** 28	0.965** 28	<b>0.920**</b> 28	0.832** 28	0.799** 28	0.893** 28
<b>T1Speech (N)</b>	- 28	- 28	- 28	- 28	1 28	0.337 28	0.825** 28	0.809** 28	<b>0.888**</b> 28	0.575** 28	0.745** 28
<b>T1Spatial (N)</b>	- 28	- 28	- 28	- 28	- 28	1 28	0.651** 28	0.658** 28	0.375* 28	<b>0.784**</b> 28	0.671** 28
<b>T1Qualities (N)</b>	- 28	- 28	- 28	- 28	- 28	- 28	1 28	0.892** 28	0.785** 28	0.767** 28	<b>0.892**</b> 28
<b>T2Total (N)</b>	- 28	- 28	- 28	- 28	- 28	- 28	- 28	1 28	0.885** 28	0.908** 28	0.957** 28
<b>T2Speech (N)</b>	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	1 28	0.647** 28	0.765** 28
<b>T2Spatial (N)</b>	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	1 28	0.871** 28
<b>T2Qualities (N)</b>	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	- 28	1 28

Note. \* correlation is significant at  $p < 0.05$  level (2-tailed). \*\* correlation is significant at the 0.01 level (2-tailed). T0 = administration one. T1 = administration two. T2 = administration three.

### 3.5 Confidence intervals and critical difference scores

The upper and lower limits of the 95% confidence intervals for total and sub-scale scores from T0, T1, and T2 data, and the critical difference scores (the range between the upper and lower limits of the 95% confidence intervals) are shown in Table 17 below.

Table 17: Lower and upper limits of the 95% confidence intervals and critical difference scores for total and sub-scale scores from T0, T1, and T2 data.

	T0 to T1		T1 to T2		T0 to T2	
	<i>Lower and upper limits of 95% CI</i>	<i>CD score</i>	<i>Lower and upper limits of 95% CI</i>	<i>CD score</i>	<i>Lower and upper limits of 95% CI</i>	<i>CD score</i>
<b>Speech</b>	0.026 – 0.474	0.448	0.021 – 0.139	0.223	-0.121 – 0.461	0.581
<b>Spatial</b>	0.391 – 0.649	0.256	0.466 – 0.734	0.269	-0.046 – 0.186	0.233
<b>Qualities</b>	0.217 – 0.463	0.247	0.202 – 0.318	0.116	-0.072 – 0.152	0.224
<b>Total</b>	0.320 – 0.521	0.201	0.232 – 0.309	0.077	-0.003 – 0.223	0.226

CI = confidence interval. CD = critical difference. T0 = administration one. T1 = administration two. T2 = administration three.

## **Chapter 4: Discussion**

### **4.1 Research question**

The aim of this research was to answer the question, “Is the SSQ-12 a reliable questionnaire for measuring hearing-aid outcomes for hearing-aid wearers when using the pen-and-paper administration method across 3 administration times: T0, T1, and T2?” The results of this study suggest that the SSQ-12 is a reliable questionnaire for measuring hearing-aid outcomes for experienced hearing-aid wearers using the pen-and-paper administration method. There were no significant differences between the SSQ-12 scores for each participant’s T0, T1, and T2 data for the total and sub-scale scores.

### **4.2 Comparison with the literature**

#### **4.2.1 Comparison with test-retest reliability data from other hearing related questionnaires.**

Most other studies reporting on the test-retest reliability of questionnaires related to hearing have calculated the test-retest reliability using correlations rather than analysis of variance. In order to readily compare these results with the test-retest reliability data from other hearing related questionnaires, test-retest reliability was calculated using Pearson’s correlations as well as analysis of variance. Test-retest reliability for the HHIE, HHIA, and HHIA-S versions ranged from  $r = 0.84 - 0.97$  for the total score (Newman et al., 1991; Weinstein et al., 1986). Cox and Alexander (1995) calculated test-retest reliability of the APHAB sub-scale scores, and values ranged from  $r = 0.65 - 0.89$ , while test-retest reliability for the PHAB sub-scale scores ranged from  $0.42 - 0.72$  (Cox & Rivera, 1992). Similar values have been reported for

other hearing related questionnaires including the SADL ( $r = 0.81$ , Cox & Alexander, 1999) and the HAPI ( $r = 0.80$ , Shum, 1993). In comparison, the test-retest reliability correlations found in this study on the SSQ-12 ( $r = 0.44 - 0.92$ ) are similar to those reported in the studies mentioned above.

Demorest and Walden (1984) recommend adequate test-retest correlations of  $r > 0.80$ . Streiner and Norman (1995) reported test-retest reliability correlations around  $r = 0.85 - 0.94$  as being acceptable. However, Streiner and Norman (1995) advised against arbitrary cut-off values, stating that sample size and population should be considered in determining what is acceptable and what is not, and that several studies should contribute to the overall test-retest reliability of the measure in question. Streiner and Norman (1995) also stated that comparing test-retest reliability to that of other self-report measures assumed to have acceptable test-retest reliability was a viable option in assessing whether test-retest reliability values are acceptable. Two parties in North America, and one in Christchurch, New Zealand are currently carrying out similar test-retest reliability studies of the SSQ-12. Data from the different sites will be combined and contribute to the overall test-retest reliability of the SSQ-12. Most of the test-retest reliability correlations found in this study would not meet the criteria of  $r = 0.85 - 0.92$ , but as pointed out, the results from this study are consistent with test-retest reliability correlations reported for other hearing related questionnaires assumed to have acceptable test-retest reliability, thus validating these results. Interestingly, poorer test-retest reliability values were found for the speech sub-scale using the T0 data (T0 – T1,  $r = 0.571$  and T0 – T2,  $r = 0.444$ ), while the T1 – T2 speech sub-scale test-retest reliability was higher ( $r = 0.888$ ) and consistent with the values found for the other sub-scales. These results show that there was more

variability in the answers given at T0 – T1 and T0 – T2, than at T1 – T2. It is not clear why this occurred.

#### **4.2.2 Comparison with test-retest reliability data from the SSQ-49.**

Overall the test-retest reliability correlations in this study are consistent with those reported by Singh and Pichora-Fuller (2010) for the SSQ-49. Singh and Pichora-Fuller (2010) reported correlations ranging from  $r = 0.56 - 0.83$  across the total and sub-scale scores for the mail-mail group, while results from this study showed correlations ranging from  $r = 0.44 - 0.92$  across the total and sub-scale scores, for T0 – T1, T0 – T2, and T1 – T2. In general, the test-retest correlations for the SSQ-12 reported in this study are higher than those reported for the mail-mail group for the SSQ-49. The higher test-retest correlations for the SSQ-12 may be related to the length of time between administrations. The SSQ-12 was administered with 6 weeks in between administrations whereas the SSQ-49 was administered with 6 months in between administrations. Lower test-retest correlations for administrations further apart in time is consistent with the literature (Streiner & Norman, 1995).

Singh and Pichora-Fuller (2010), found lower test-retest reliability correlations for the spatial sub-scale ( $r = 0.49 - 0.59$ ) for all the administration methods apart from the interview-interview administration group ( $r = 0.86$ ). This study found lower test-retest correlations for the speech sub-scale ( $r = 0.44 - 0.57$ ) for the T0 – T2 and T0 – T1 correlations, respectively, but not for the T1 – T2 speech sub-scale correlation. Singh and Pichora-Fuller (2010) attributed the lower test-retest correlations in the spatial sub-scale for all the groups except the interview-interview group, to the fact that participants could not clarify understanding of the context or meaning of the questions. They suggested that poorer understanding of the questions lead to less reliable answers in the spatial sub-scale. Results from the present study do not support

their conclusion. Despite using the pen-and-paper administration method (mail), the results from the present study of the SSQ-12 test-retest reliability did not show a similar pattern of lower test-retest reliability for the spatial sub-scale.

In contrast to the SSQ-49 results, results from the present study showed lower test-retest correlations on the speech sub-scale for the T0 – T1, and T0 – T2 administration periods. One explanation for this could be that the instructions were not clear as to how participants should complete the SSQ-12, i.e., how they hear with their hearing aids, or how they hear without their hearing aids. Some participants expressed confusion or uncertainty about this when they completed the SSQ-12 the first time. Given the lower test-retest correlation is only present where the T0 data is included, it may be that some participants completed the SSQ-12 at T0 from the perspective of how they hear without their hearing aids, and the T1 and T2 data as how they hear with their hearing aids. This could explain why the T0 – T1, and T0 – T2 administration periods showed lower test-retest correlations, while the T1 – T2 test-retest correlations showed no significant differences between the speech and other sub-scale or total score. However, this explanation does not explain why it only affected the speech sub-scale and not the other sub-scales too. Another explanation could be that the speech sub-scale is more prone to variability based on participants' most recent experiences of trying to understand speech in difficult contexts. Given the importance of understanding speech to one's ability to communicate in everyday life, it is plausible that participants would pay more attention to their ability to understand speech than telling where a sound is coming from, or noting the quality of a sound, which may explain why the speech sub-scale was affected, while the other sub-scales were not. Still this does not explain why it affected T0 – T1, and T0 – T2, and not T1 – T2 for the speech sub-scale.

### **4.3 Critical difference scores**

Demorest and Walden (1984) and Hyde (2000) recommend when assessing test-retest reliability studies to also report critical difference scores. The value of reporting critical difference scores is that clinicians can readily assess whether a change in score following treatment or intervention is significant or not, whereas with only test-retest correlation data it is not clear what those values translate to in terms of significant changes pre and post an intervention (Cox & Alexander, 1999).

#### **4.3.1 Comparison with critical difference scores from other hearing related questionnaires.**

Several test-retest reliability studies for hearing related questionnaires have reported critical difference scores. Weinstein et al. (1986) reported relatively large critical difference scores for the HHIE when using pen-and-paper administration (+/- 36%) and interview method of administration (+/- 18.7%). Similar critical difference scores have been reported by Cox and Alexander (1995) for the APHAB (+/- 21 - 37%), Cox and Gilmore (1990) for the PHAP (+/- 18 - 28%), and Cox and Rivera (1992) for the PHAB (+/- 25 - 32%). Newman et al. (1991) reported critical difference scores of +/- 11.9% for the HHIA, and +/- 22% for the HHIA-S. In comparison, the present study calculated significantly lower critical difference scores (0.77% – 2.26%) indicating the SSQ-12 is more sensitive to change than the questionnaires mentioned above.

#### **4.3.2 Comparison with critical difference scores from the SSQ-49.**

Singh and Pichora-Fuller (2010) reported critical difference scores of 0.4 points (4%) and 0.6 points (6%) for the interview and pen-and-paper administration methods, respectively. The critical difference scores of the SSQ-12 from the present



study are similar to those reported by Singh and Pichora-Fuller (2010), but are lower ranging from 0.077 points (0.77%) for T1 – T2, to 0.226 points (2.26%) for T0 – T2.

#### **4.4 Research limitations**

Limitations of this research include inadequate sampling and subject self-selection bias. No randomisation was used for recruitment. Participants were selected by the researcher from a clinic database with the intent of recruiting a gender balanced sample with a wide range of ages, but this method of recruitment risks experimenter selection bias resulting in inadequate sampling (i.e., the sample is not representative of the population of interest; Cox, 2005).

Subject self-selection bias is an inherent limitation in most research, whereby those who do not want to participate in research vary in characteristics or variables from those who readily volunteer to participate in research studies. Those who opt out of participating in research studies may have certain characteristics or variables that influence the outcome of a study, but are not represented because they are not inclined to participate. While the participation rate in this study was high, most of the people who participated in the study had opted to be contacted for future research studies, and therefore may have certain characteristics that people who did not opt to be contacted for future research do not have, leading to inadequate sampling of the population or bias in the results.

Overall, the sample recruited represented a wide range of characteristic variables including age, gender, pure-tone average, hearing-aid use time, income bracket, and level of education, so the results are generalisable to a wider population. The sample was made up of mainly New Zealand European participants. However, the study is also being conducted in different locations around the world, which will

reduce the possible bias introduced in this study with a predominance of New Zealand European participants.

Another limitation of this research is the lack of blinding of the participants. They were aware of what the study was looking for, and this may have impacted their responses on the questionnaire. A limitation of the questionnaire (SSQ-12) used in the study, is that it did not state clearly how participants should respond to the questions, i.e., to respond based on how they hear with their hearing aids, or based on how they hear without their hearing aids. Some participants expressed confusion over this matter, which may have confounded their responses on the SSQ-12 and the results of the test-retest reliability, as explained earlier. The SSQ-12 comparison (SSQ-12-C) and benefit (SSQ-12-B) versions state clearly how they are to answer the questions, i.e., based on how they hear with their hearing aids (SSQ-12-B), or compared to their previous hearing aids (SSQ-12-C; MRC Institute of Hearing Research, n.d.). In future test-retest reliability or validity studies of questionnaires, it should be made clear to the respondents how they are to complete it (i.e., with or without hearing aids) to avoid confounding the results.

#### **4.5 Future research**

The present research confirms that the SSQ-12 is a reliable self-report tool for hearing-aid users, using the pen-and-paper administration method. Previous research has shown that test-retest reliability can vary on the same questionnaire depending on how it is administered, i.e., self-administration (pen-and-paper) or interview (face-to-face). Future research needs to assess the test-retest reliability of the SSQ-12 using interview administration, and combinations of the two administration methods. Presently there is only one published study that has used the SSQ-12 as an outcome measure. In the future, more studies need to be carried out using the SSQ-12 as an

outcome measure, so that the validity of the SSQ-12 as an outcome measure can be evaluated.

#### **4.6 Clinical implications**

The results of this research show that the SSQ-12 is a reliable self-report measure for hearing-aid users using a self-administration (pen-and-paper) method, when administered approximately 6 weeks apart. This means that clinicians can use the SSQ-12 as a benefit measure (i.e., administer it before and after an intervention or treatment) and be confident that a change in mean score equal to or greater than the critical difference scores calculated in this study is due to the intervention and not by chance.

#### **4.7 Summary**

The aim of this research was to investigate if the SSQ-12 is a reliable questionnaire for measuring hearing-aid outcomes for hearing-aid users using the pen-and-paper administration method. Twenty-eight experienced hearing-aid wearers were recruited from a clinic database and administered the SSQ-12 questionnaire 3 times, with 6 weeks in between administrations. A repeated-measures ANOVA was used to analyse responses across the three administration times. The results showed no statistically significant differences between each participant's responses across the three administration times (T0, T1, and T2), indicating that the SSQ-12 is a reliable questionnaire for measure hearing-aid outcomes for experienced hearing-aid wearers using the pen-and-paper administration method. Critical difference scores were calculated to facilitate clinicians using the SSQ-12 as a hearing-aid outcome measure in the disability domain, by stating how much of a change in score constitutes a significant change. It is important to know that the SSQ-12 questionnaire is a reliable hearing-aid outcome measure, so that clinicians can be confident that a change in

score on the SSQ-12 greater than or equal to the critical difference scores reported in this study is related to the intervention or treatment a patient receives rather than measurement error.

## References

- Accident Compensation Corporation. (2014). *ACC611 Hearing aid trial outcome report*. Retrieved from <http://www.acc.co.nz/publications/index.htm?ssBrowseSubCategory=Audiology>
- Akeroyd, M. A., Guy, F. H., Harrison, D. L., & Suller, S. L. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). *International Journal of Audiology*, 53(2), 101-114.
- American Speech-Language-Hearing Association (ASHA). (2006). *Preferred practice patterns for the profession of audiology* [Preferred Practice Patterns]. Retrieved from <http://www.asha.org/policy/PP2006-00274/#sec1.4.14>
- Anderson, K., & Smaldino, J. (1996). *Listening inventories for education (L.I.F.E.): An efficacy tool*. Tampa, FL: Educational Audiology Association.
- Arndt, S., Aschendorff, A., Laszig, R., Beck, R., Schild, C., Kroeger, S., . . . Wesarg, T. (2010). Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otology & Neurology*, 32, 39-47.
- Arndt, S., Prose, S., Laszig, R., Wesarg, T., Aschendorff, A., & Hassepass, F. (2015). Cochlear implantation in children with single-sided deafness: Does aetiology and duration of deafness matter? *Audiology & Neurotology*, 20(supplement 1), 21-30.

- Banh, J., Singh, G., & Pichora-Fuller, M. K. (2012). Age affects responses on the speech, spatial, and qualities of hearing scale (SSQ) by adults with minimal audiometric loss. *Journal of the American Academy of Audiology*, 23(2), 81-91.
- Beijen, J. W., Mylanus, E. A. M., & Snik, A. F. M. (2007). Sound localization ability of young children with bilateral cochlear implants. *Otology & Neurology*, 28(4), 479-485.
- Bentler, R. A., & Kramer, S. E. (2000). Guidelines for choosing a self-report outcome measure. *Ear and Hearing*, 21(4), 37S-49S.
- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organisation of sound. Cambridge, MA: MIT Press.
- British Society of Audiology. (2012). *Practice guidance: Common principles of rehabilitation for adults with hearing and/or balance-related problems in routine audiology services*. Retrieved from [http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA\\_PPC\\_Rehab\\_Final\\_30August2012.pdf](http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_PPC_Rehab_Final_30August2012.pdf)
- Capretta, N. R., & Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *The Laryngoscope*, 126, 699-706.
- Ching T.C., Hill M., Psarros C. Strategies for evaluation of hearing aid fitting for children. International Hearing Aid Research Conference. Lake Tahoe, USA; 2000.

- Ching, T. Y. C., & Hill, M. (2007). The parents' evaluation of aural/oral performance of children (PEACH) scale: Normative data. *Journal of the American Academy of Audiology*, 18(3), 220-235.
- Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical Interventions in Aging*, 20, 159-163.
- Coninx F., Weichbold V., Tsiakpini L. *LittleEARS® Auditory Questionnaire*. MED-EL, 2003.
- Cox, R. M., Hyde, M. L., Gatehouse, S., Noble, W., Dillon, H., Bentler, R. A., . . . Hallberg, L. (2000). Optimal outcome measures, research priorities, and international cooperation. *Ear and Hearing*, 21(4), 106S-115S.
- Cox, R. M. (2003). Assessment of subjective outcome of hearing aid fitting: getting the client's point of view. *International Journal of Audiology*, 42, 90-96.
- Cox, R. M. (2005). Choosing a self-report measure for hearing aid fitting outcomes. *Seminars in Hearing*, 26(3), 149-156.
- Cox, R. M., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, 16(2), 176-186.
- Cox, R. M., & Alexander, G. C. (1999). Measuring satisfaction with amplification in daily life: the SADL scale. *Ear and Hearing*, 20(4), 306-319.
- Cox, R. M., & Gilmore, C. (1990). Development of the Profile of Hearing Aid Performance (PHAP). *Journal of Speech and Hearing Research*, 33(2), 343-357.

- Cox, R. M., Johnson, J. A., & Xu, J. (2014). Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. *Gerontology*, 60(6), 557-568.
- Cox, R. M., & Rivera, I. M. (1992). Predictability and reliability of hearing aid benefit measured using the PHAB. *Journal of the American Academy of Audiology*, 3(4), 242-254.
- Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *The Gerontologist*, 43(5), 661-668.
- Dawes, P., Munro, K. J., Kalluri, S., & Edwards, B. (2013). Acclimatization to hearing aids. *Ear and Hearing*, 35(2), 203-212.
- Demeester, K., Topsakal, V., Hendrickx, J., Fransen, E., van Laer, L., Van Camp, G., . . . van Wieringen, A. (2012). Hearing disability measured by the speech, spatial, and qualities of hearing scale in clinically normal-hearing and hearing-impaired middle-aged persons, and disability screening by means of a reduced SSQ (the SSQ5). *Ear and Hearing*, 33(5), 615-626.
- Demorest, M. E., & DeHaven, G. P. (1993). Psychometric adequacy of self-assessment scales. *Seminars in Hearing*, 14(4), 314-325.
- Demorest, M. E., & Erdman, S. A. (1987). Development of the communication profile for the hearing impaired. *Journal of Speech and Hearing Disorders*, 52, 129-143.
- Demorest, M. E., & Erdman, S. A. (1988). Retest stability of the communication profile for the hearing impaired. *Ear and Hearing*, 9(5), 237-242.



- Demorest, M. E., & Walden, B. E. (1984). Psychometric Principles in the Selection, Interpretation, and Evaluation of Communication Self-Assessment Inventories. *Journal of Speech and Hearing Disorders*, 49, 226-240.
- Dillon, H., James, A., & Ginis, J. (1997). Client oriented scale of improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. *Journal of the American Academy of Audiology*, 8, 27-43.
- Douglas, S. A., Yeung, P., Daudia, A., Gatehouse, S., & O'Donoghue, G. M. (2007). Spatial hearing disability after acoustic neuroma removal. *The Laryngoscope*, 117, 1648-1651.
- Dwyer, N. Y., Firszt, J. B., & Reeder, R. M. (2014). Effects of unilateral input and mode of hearing in the better ear: Self-reported performance using the speech, spatial and qualities of hearing scale. *Ear and Hearing*, 35(1), 126-136.
- Ellis, R. J., & Munro, K. J. (2015). Benefit from, and acclimatization to, frequency compression hearing aids in experienced adult hearing-aid users. *International Journal of Audiology*, 54(1), 37-47.
- Emmett, S. D., & Francis, H. W. (2014). The socioeconomic impact of hearing loss in U.S. adults. *Otology & Neurotology*, 36, 545-550.
- Erdman, S. A. (1993). Self-assessment in audiology: The clinical rationale. *Seminars in Hearing*, 14(4), 303-313.
- Fang, T., Chang, H., Wan, T., Wang, P., & Chen, Y. (2013). Validation of the Chinese version "satisfaction with amplification in daily life (SADL)" survey for hearing-aid users. *International Journal of Audiology*, 52(7), 478-484.

- Farinetti, A., Roman, S., Mancini, J., Baumstarck-Barrau, K., Meller, R., Lavieille, J. P., & Triglia, J. M. (2015). Quality of life in bimodal hearing users (unilateral cochlear implants and contralateral hearing aids). *European Archives of Oto-Rhino-Laryngology*, 272, 3209-3215.
- Flexer, C., & Madell, J. R. (2014). Why hearing is important in children. In C. Flexer & J. R. Madell (Eds.), *Pediatric Audiology: Diagnosis, Technology, and Management* (2nd ed., pp. 3-7). New York: Thieme.
- Froman, R. D. (2014). The ins and outs of self-report response options and scales. *Research in Nursing & Health*, 37, 447-451.
- Galvin, K. L., & Mok, M. (2016). Everyday listening performance of children before and after receiving a second cochlear implant: results using the parent version of the speech, spatial, and qualities of hearing scale. *Ear and Hearing*, 37(1), 93-102.
- Galvin, K. L., & Noble, W. (2013). Adaptation of the speech, spatial, and qualities of hearing scale for use with children, parents, and teachers. *Cochlear Implants International*, 14(3), 135-141.
- Galvin, K. L., Sarant, J. Z., Harris, D. C., Bennet, L. A., Mok, M., & Canagasabey, M. (2015). Performance ratings for children using bilateral cochlear implants obtained with the speech spatial, and other qualities of hearing scale for parents. *Cochlear Implants International*, 16(S1), S19-S22.
- Gatehouse, S. (1999). Glasgow hearing aid benefit profile: derivation and validation of a client-centered outcome measure for hearing aid services. *Journal of the American Academy of Audiology*, 10, 80-103.

- Gatehouse, S., & Akeroyd, M. A. (2007). *Tailoring hearing-aid and rehabilitative interventions to listeners' disabilities and priorities*. Paper presented at the Hearing Care for Adults: An International Conference, Chicago, IL.
- Gatehouse, S., Elberling, C., & Naylor, G. (1999). Aspects of auditory ecology and psychoacoustic function as determinants of benefits from and candidature for non-linear processing in hearing aids. In Auditory models and non-linear hearing instruments, *18<sup>th</sup> Danaavox Symposium*, 221-233.
- Gatehouse, S., Naylor, G., & Elberling, C. (2006b). Linear and nonlinear hearing aid fittings - 2. Patterns of candidature. *International Journal of Audiology*, *45*(3), 153-171.
- Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*(2), 85-99.
- Goldfried, M. R., & D'Zurilla, T. J. (1969). A behavioural-analytic method for assessing competence. In C. D. Spielberger (Ed.), *Current Topics in Clinical and Community Psychology*. New York: Academic.
- Hartman, D. P., Roper, B. L., & Bradford, D. C. (1979). Some relationships between behavioural and traditional assessment. *Journal of Behavioural Assessment*, *1*, 3-21.
- Hassepass, F., Aschendorff, A., Wesarg, T., Kröger, S., Laszig, R., Beck, R., . . . Arndt, S. (2013). Unilateral deafness in children: Audiologic and subjective assessment of hearing ability after cochlear implantation. *Otology & Neurology*, *34*(1), 53-60.

- Hassepass, F., Schild, C., Aschendorff, A., Laszig, R., Maier, W., Beck, R., . . . Arndt, S. (2013). Clinical outcome after cochlear implantation in patients with unilateral hearing loss due to labyrinthitis ossificans. *Otology & Neurology*, 34, 1278-1283.
- Heine, C., & Browning, C. J. (2002). Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions. *Disability and Rehabilitation*, 24(15), 763-773.
- Heo, J., Lee, J., & Lee, W. (2013). Bimodal benefits on objective and subjective outcomes for adult cochlear implant users. *Korean Journal of Audiology*, 17(2), 65-73.
- Héту, R., Getty, L., Philibert, L., Desilets, F., Noble, W., et al. 1994. Mise au point d'un outil clinique pour la mesure d'incapacités auditives et de handicaps (Development of a clinical tool for the measurement of the severity of hearing disabilities and handicaps). *Journal of Speech & Language Pathology & Audiology*, 18, 83–95.
- High, W., Fairbanks, G., & Glorig, A. (1964). Scale for self-assessment of hearing handicap. *Journal of Speech and Hearing Disorders*, 29(3), 215-230.
- Hinderink, J. B., Krabbe, P. F. M., & Van den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The Nijmegen Cochlear Implant Questionnaire. *Otolaryngology - Head and Neck Surgery*, 123, 756-765.
- House, J., Kutz, J., Chung, J., & Fisher, L. (2010). Bone-anchored hearing aid subjective benefit for unilateral deafness. *The Laryngoscope*, 120, 601-607.

- Hyde, M. L. (2000). Reasonable psychometric standards for self-report outcome measures in audiological rehabilitation. *Ear and Hearing, 21*(4), 24S-36S.
- Jensen N.S., Akeroyd M.A., Noble W. & Naylor G. (2009). The Speech, Spatial and Qualities of Hearing scale (SSQ) as a benefit measure. Poster presented at the NCRAR conference on “The Ear-Brain System: Approaches to the Study and Treatment of Hearing Loss”, Portland, October 2009 (see <http://www.ihr.mrc.ac.uk/products/display/ssq>).
- Jensen, B. J., & Haynes, S. N. (1986). Self-report questionnaires and inventories. In A. R. Ciminero, S. C. Karen, & H. E. Adams (Eds.), *Handbook of Behavioural Assessment* (2 ed., pp. 150-175). New York: John Wiley & Sons, Inc.
- Jespersen, C. T., Bille, M., & Legarth, J. V. (2014). Psychometric properties of a revised Danish translation of the international outcome inventory for hearing aids (IOI-HA). *International Journal of Audiology, 53*(5), 302-308.
- Johnson, C. E., & Danhauer, J. L. (2002). *Handbook of Outcomes Measurement in Audiology*. Canada: Thomson Delmar Learning.
- Kam, A. C. S., Tong, M. C. F., & van Hasselt, A. (2011). Cross-cultural adaptation and validation of the Chinese abbreviated profile of hearing aid benefit. *International Journal of Audiology, 50*(5), 334-339.
- Kamil, R. J., & Lin, F. R. (2015). The effects of hearing impairment in older adults on communication partners: A systematic review. *Journal of the American Academy of Audiology, 26*, 155-182.
- Kiessling, J., Grugel, L., Meister, H., & Meis, M. (2011). German translations of

- questionnaires SADL, ECHO, and SSQ and their evaluation. *Zeitschrift für Audiologie*, 50, 6-16.
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Tobi, H. (1995). Factors in subjective hearing disability. *Audiology*, 34(6), 311-320.
- Laske, R. D., Rösli, C., Pfiffner, F., Veraguth, D., & Huber, A. M. (2015). Functional results and subjective benefit of a transcutaneous bone conduction device in patients with single-sided deafness. *Otology & Neurology*, 36(7), 1151-1156.
- Lovett, R. E., Kitterick, P. T., Hewitt, C. E., & Summerfield, A. Q. (2010). Bilateral or unilateral cochlear implantation for deaf children: an observational study. *Archives of Disease in Childhood*, 95(2), 107-112.
- Martin, T. P. C., Lowther, R., Cooper, H., Holder, R. L., Irving, R. M., Reid, A. P., & Proops, D. (2010). The bone-anchored hearing aid in the rehabilitation of single-sided deafness: experience with 58 patients. *Clinical Otolaryngology*, 35, 284-290.
- McCreery, R. W., Walker, E. A., Spratford, M., Oleson, J., Bentler, R. A., Holte, L., & Roush, P. (2015). Speech recognition and parent ratings from auditory development questionnaires in children who are hard of hearing. *Ear and Hearing*, 36, 60S-75S.
- McGrath, M. C. (2011). Psychometrics. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behaviour and Development* (pp. 1190-1191). Online: Springer. Retrieved from [http://gr2tq4rz9x.search.serialssolutions.com/?ctx\\_ver=Z39.88-](http://gr2tq4rz9x.search.serialssolutions.com/?ctx_ver=Z39.88-)

2004&ctx\_enc=info%3Aofi%2Fenc%3AUTF-  
8&rft\_id=info:sid/summon.serialssolutions.com&rft\_val\_fmt=info:ofi/fmt:ke  
v:mtx:book&rft.genre=bookitem&rft.title=Encyclopedia+of+Child+Behavior  
+and+Development&rft.au=McGrath%2C+Marie&rft.atitle=Psychometrics&r  
ft.date=2011-01-  
01&rft.isbn=9780387790619&rft.spage=1190&rft.epage=1191&rft.externalD  
ocID=3707000875&paramdict=en-US

McNeil, M. L., Gulliver, M., Morris, D. P., Makki, F. M., & Bance, M. (2014). Can audiometric results predict qualitative hearing improvements in bone-achored hearing aid recipients? *The Journal of Laryngology & Otology*, 128, 35-42.

Mertens, G., Punte, A. K., & Van de Heyning, P. (2013). Self-assessment of hearing disabilities in cochlear implant users using the SSQ and the reduced SSQ5 version. *Otology & Neurology*, 34, 1622-1629.

Mertens, G., & Van de Heyning, P. (2016). Letter to the editor. *Otology & Neurology*, 37, 813.

Montano, J. J. (2015). Audiologic rehabilitation. In J. Katz, M. Chasin, K. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (7th ed., pp. 849-860). Philadelphia: Wolters Kluwer.

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing* (6<sup>th</sup> Ed.). Leiden: Brill.

Moulin, A., & Richard, C. (2016). Sources of variability of speech, spatial, and qualities of hearing scale (SSQ) scores in normal-hearing and hearing-

impaired populations. *International Journal of Audiology*, 55(2), 101-109.

doi:10.3109/14992027.2015.1104734

MRC Institute of Hearing Research. (n.d.). The speech, spatial and qualities of hearing scale (SSQ). Retrieved from <https://www.ihr.mrc.ac.uk/pages/products/ssq>

Musiek, F. E., & Baran, J. A. (2007). *The Auditory System: Anatomy, Physiology, and Clinical Correlates*. Boston, MA: Pearson Education, Inc.

Newman, C. W., & Weinstein, B. E. (1989). Test-retest reliability of the hearing handicap inventory for the elderly using two administration approaches. *Ear and Hearing*, 10(3), 190-191.

Newman, C. W., Weinstein, B. E., Jacobson, G. P., & Hug, G. A. (1990). The hearing handicap inventory for adults: Psychometric adequacy and audiometric correlates. *Ear and Hearing*, 11(6), 430-433.

Newman, C. W., Weinstein, B. E., Jacobson, G. P., & Hug, G. A. (1991). Test-retest reliability of the hearing handicap inventory for adults. *Ear and Hearing*, 12(5), 355-357.

New Zealand Audiological Society (2016). *NZAS best practice guidelines adult hearing aid fitting: abbreviated best practice guidelines*. Auckland, New Zealand: Author.

Noble, W. (1978). *Assessment of Impaired Hearing: a Critique and a New Method*. New York: Academic Press.



- Noble, W. (1993). Preface: An historical perspective on self-assessment in audiology. *Seminars in Hearing, 14*(4), 299-302.
- Noble, W. (1998). *Self-Assessment of Hearing and Related Functions*. London: Whurr Publishers Ltd.
- Noble, W. (2010). Assessing binaural hearing: Results using the speech, spatial and qualities of hearing scale. *Journal of the American Academy of Audiology, 21*(9), 586-574.
- Noble, W. (2013). *Self-Assessment of Hearing* (2nd ed.). San Diego, CA: Plural Publishing, Inc.
- Noble, W., & Gatehouse, S. (2004). Interaural asymmetry of hearing loss, speech, spatial and qualities of hearing scale (SSQ) disabilities and handicap. *International Journal of Audiology, 43*(2), 100-114.
- Noble, W., & Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ). *International Journal of Audiology, 45*(3), 172-181.
- Noble, W., Jensen, N. S., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2013). A short form of the speech, spatial and qualities of hearing scale suitable for clinical use: The SSQ12. *International Journal of Audiology, 52*(6), 1-7.
- Noble, W., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2012). Self-assessed hearing abilities in middle- and older-age adults: A stratified sampling approach. *International Journal of Audiology, 51*, 174-180.

- Noble, W., Ter-Horst, K. & Byrne, D. (1995). Disabilities and handicaps associated with impaired auditory localization. *Journal of the American Academy of Audiology*, 6(2), 129-140.
- Noble, W., Tyler, R. S., Dunn, C. C., & Bhullar, N. (2008). Hearing handicap ratings among different profiles of adult cochlear implant users. *Ear and Hearing*, 29(1), 112-120.
- Noble, W., Tyler, R. S., Dunn, C. C., & Bhullar, N. (2008b). Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: Comparing self-assessed and measured abilities. *International Journal of Audiology*, 47(8), 505-514.
- Noble, W., Tyler, R. S., Dunn, C. C., & Bhullar, N. (2009). Younger- and older-age adults with unilateral and bilateral cochlear implants: speech and spatial hearing self-ratings and performance. *Otology & Neurology*, 30, 921-929.
- Nunnally J.C. 1978. *Psychometric Theory (2nd Ed.)*. New York: McGraw Hill.
- Olsen, S. Ø., Hernvig, L. H., & Nielson, L. H. (2012). Self-reported hearing performance among subjects with unilateral sensorineural hearing loss. *Audiological Medicine*, 10, 83-92.
- Pai, I., Kelleher, C., Nunn, T., Pathak, N., Jindal, M., O'Connor, A., & Jiang, D. (2012). Outcome of bone-anchored hearing aids for single-sided deafness: A prospective study. *Acta Oto-Laryngologica*, 132(7), 751-755.
- Perez, E., McCormack, A., & Edmonds, B. A. (2014). Sensitivity to temporal fine structure and hearing-aid outcomes in older adults. *Frontiers in Neuroscience*, 8, 1-9.

- Potts, L. G., & Litovsky, R. Y. (2014). Transitioning from bimodal to bilateral cochlear implant listening: Speech recognition and localization in four individuals. *American Journal of Audiology*, 23(1), 79-92.
- Redfors, Y. D., Olaison, S., Karlsson, J., Hellgren, J., & Möller, C. (2015). Hearing-related, health-related quality of life in patients who have undergone otosclerosis surgery: A long-term follow-up study. *International Journal of Audiology*, 54(2), 63-69.
- Robinson, K., Gatehouse, S. & Browning, G.G. 1996. Measuring patient benefit from otorhinolaryngological surgery and therapy. *The Annals of Otology, Rhinology & Laryngology*, 105(6), 415–422.
- Ryu, N., Moon, I. J., Byun, H., Jin, S. H., Park, H., Jang, K., & Cho, Y. (2015). Clinical effectiveness of wireless CROS (contralateral routing of offside signals) hearing aids. *European Archives of Oto-Rhino-Laryngology*, 272(9), 2213-2219.
- Schlauch, R. S., Nelson, P. (2015). Puretone evaluation. In J. Katz, M. Chasin, K. English, L. J. Hood & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (7<sup>th</sup> Ed., pp. 29-47) Philadelphia: Wolters Kluwer Health.
- Schow, R. L., & Nerbonne, M. A. (1982). Communication screening profile: use with elderly adults. *Ear and Hearing*, 3, 135-147.
- Schum, D. J. (1993). Test-retest reliability of a shortened version of the hearing aid performance inventory. *Journal of the American Academy of Audiology*, 4(1), 18-21.

- Schwartz, K., Ringleb, S. I., Sandberg, H., Raymer, A., & Watson, G. S. (2015). Development of trivia game for speech understanding in background noise. *International Journal of Speech-Language Pathology*, 17(4), 357-366.
- Singh, G., & Pichora-Fuller, M. K. (2010). Older adults' performance on the speech, spatial, and qualities of hearing scale (SSQ): test-retest reliability and a comparison of interview and self-administration methods. *International Journal of Audiology*, 49, 733-740.
- Sparreboom, M., Snik, A. F. M., & Mylanus, E. (2012). Sequential bilateral cochlear implantation in children: quality of life. *Archives of Otolaryngology - Head & Neck Surgery*, 138(2), 134-141.
- Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., . . . Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom cochlear implant system. *Ear and Hearing*, 28, 62-72.
- Stephens, D. (2001). World Health Organization's international classification of functioning, disability, and health - ICF. *Journal of Audiological Medicine*, 10(3), vii-x.
- Streiner, D. L., & Norman, G. R. (1995). *Health Measurement Scales: A Practical Guide to Their Development and Use* (2<sup>nd</sup> Ed.). Oxford: Oxford University Press.
- Summerfield, A. Q., Barton, G. R., Toner, J., McAnallen, C., Proops, D., Harries, C., . . . Pringle, M. (2006). Self reported benefits from successive bilateral cochlear

- implantation in post-lingually deafened adults: randomised controlled trial. *International Journal of Audiology*, 45(Supplement 1), S99-S107.
- Távora-Vieira, D., Marino, R., Krishnaswamy, J., Kuthbutheen, J., & Rajan, G. P. (2013). Cochlear implantation for unilateral deafness with and without tinnitus: A case series. *The Laryngoscope*, 123, 1251-1255.
- Tye-Murray, N. (2015). *Foundations of Aural Rehabilitation: Children, Adults, and Their Family Members* (4<sup>th</sup> Ed.). Stamford, CT: Cengage Learning.
- Tyler, R. S., Perreau, A. E., & Ji, H. (2009). Validation of the spatial hearing questionnaire. *Ear and Hearing*, 30, 466–474.
- Vannson, N., James, C., Fraysse, B., Strelnikov, K., Barone, P., Deguine, O., & Marx, M. (2015). Quality of life and auditory performance in adults with asymmetric hearing loss. *Audiology & Neurotology*, 20, 38-43.
- Ventry, I. M., & Weinstein, B. E. (1982). The hearing handicap inventory for the elderly: A new tool. *Ear and Hearing*, 3(3), 128-134.
- Walden, B. E., Demorest, M. E., & Hepler, E. L. (1984). Self-report approach to assessing benefit derived from amplification. *Journal of Speech and Hearing Research*, 27, 49-56.
- Weinstein, B. E., Spitzer, J. B., & Ventry, I. M. (1986). Test-retest reliability of the hearing handicap inventory for the elderly. *Ear and Hearing*, 7(5), 295-299.
- Weinstein, B. E. (1993). Validation of self-assessment scales as outcome measures in hearing aid fitting. *Seminars in Hearing*, 14(4), 326-337.

- Weinstein, B. E. (2000). Outcome measures in rehabilitative audiology. In J. G. Alpiner & P. A. McCarthy (Eds.), *Rehabilitative Audiology: Children and Adults* (3rd Ed.). Baltimore, MD: Lippincott, Williams, & Wilkins.
- Weinstein, B. E. (2015). Hearing loss in the elderly: A new look at an old problem. In J. Katz, M. Chasin, K. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (7<sup>th</sup> Ed., pp. 631-645). Philadelphia: Wolters Kluwer.
- Weinstein, B. E., & Ventry, I. M. (1982). Hearing impairment and social isolation in the elderly. *Journal of Speech and Hearing Research*, 25, 593-599.
- Weinstein, B. E., & Ventry, I. M. (1983). Audiometric correlates of the hearing handicap inventory for the elderly. *Journal of Speech and Hearing Disorders*, 48, 379-384.
- Winn, S. (2007). Employment outcomes for people in Australia who are congenitally deaf: Has anything changed? *American Annals of the Deaf*, 152(4), 382-390.
- World Health Organization (1980). *International Classification of Impairments, Disabilities, and Handicaps*. Geneva, Switzerland: Author.
- World Health Organization (2001). *International Classification of Functioning, Disability and Health*. Geneva, Switzerland: Author.
- World Health Organization (2002). *Towards a Common Language for Functioning, Disability, and Health: ICF*. Geneva, Switzerland: Author.
- World Health Organization (WHO). 2016. *Prevention of Blindness and Deafness: Estimates*. Retrieved from <http://www.who.int/pbd/deafness/estimates/en/>

World Health Organization (WHO). 2015. *Deafness and Hearing Loss* (Fact sheet No. 300). Retrieved from

<http://www.who.int/mediacentre/factsheets/fs300/en/>

Zahorik, P., & Rothpletz, A. M. (2015). Speech, spatial, and qualities of hearing scale (SSQ): Normative data from young, normal-hearing listeners. *The Journal of the Acoustical Society of America*, 135(4), 2163-2164.

Zhang, J., Tyler, R., Ji, H., Dunn, C., Wang, N., Hansen, M., & Gantz, B. (2015). Speech, spatial and qualities of hearing scale (SSQ) and spatial hearing questionnaire (SHQ) changes over time in adults with simultaneous cochlear implants. *American Journal of Audiology*, 24(3), 384-397.

# Appendices

## Appendix A: Ethics approval letter.



HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen  
Email: [human-ethics@canterbury.ac.nz](mailto:human-ethics@canterbury.ac.nz)

Ref: HEC 2015/85

5 August 2015

Rebecca Kelly-Campbell  
Department of Communication Disorders  
UNIVERSITY OF CANTERBURY

Dear Rebecca

The Human Ethics Committee advises that your research proposal "Test-retest reliability of the speech, spatial and qualities of hearing scale - short form (SSQ12): comparison of administration method and hearing aid experience" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 4 August 2015.

Best wishes for your project.

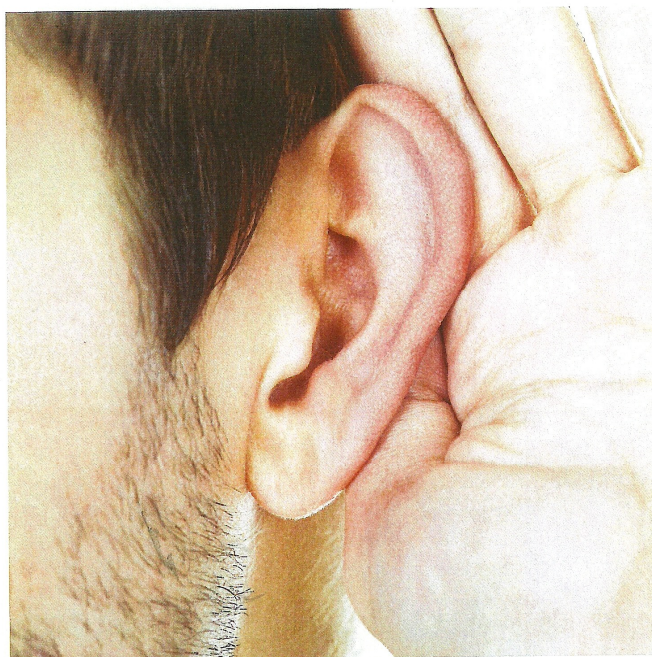
Yours sincerely

A handwritten signature in black ink, appearing to read 'L. MacDonald'.

Lindsey MacDonald  
**Chair**  
*University of Canterbury Human Ethics Committee*



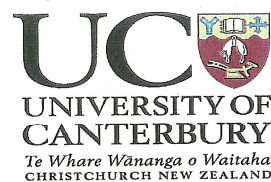
## Appendix B: Poster advertisement.



**ARE YOU AN ADULT  
WITH HEARING LOSS?  
WE NEED YOUR HELP!**

One of the most common problems people with hearing loss have is trouble understanding speech in difficult situations (like noise). A new survey was developed to measure people's experiences in those situations. This study aims to test the reliability of that survey to make sure it is useful in a clinical setting.

To take part in this research or for more information, please contact:



**You need to be:**

OR

**You will:**

You will be randomly selected to complete them at home or at the University of Canterbury

You will receive a \$10 petrol voucher if you are selected complete them at the university.

If you don't have one, we will test your hearing at no cost to you.

[illegible]

## Appendix C: Research information and consent forms.



### Research Information Sheet

**Study Title:** Test-retest reliability of the *Speech, Spatial, and Qualities of Hearing Scale – Short Form (SSQ12)*: Comparison of administration method and hearing aid experience

**Researcher:**

Dr Rebecca Kelly-Campbell  
Senior Lecturer  
Department of Communication Disorders  
University of Canterbury  
Email: rebecca.kelly@canterbury.ac.nz  
Phone: (03) 364 2987 ext 8327

**What is the aim of the study?**

- To assess the reliability of a survey used in hearing clinics (*SSQ12: Speech, Spatial, and Qualities of Hearing Scale – Short Form*)
- To test for any differences on the survey between people who have never used hearing aids and people who wear hearing aids.

**Who do you need for the study?**

I need 2 groups of people:

- Adults who have never used hearing aids
- Adults who do wear hearing aids

**What will happen in the study?**

You will be asked to complete surveys about you and your hearing. You will be asked to complete the SSQ12 three times, each around 6 weeks apart. The first time you complete the surveys, it will take you about 20 minutes. The second and third times, you only have one short survey to complete so it should take less than 10 minutes of your time.

If you have had a recent hearing test, I will also ask your permission to get those results. If you have not had a recent hearing test or do not want to release them, I can check your hearing at no cost to you. During that hearing check, you will hear several tones. I will ask you to push a button every time you hear the tone. The hearing check will take about 20 minutes.

People who sign up for the study will be randomly assigned to a group. One group will complete the survey through the post. The other group will complete the survey through an interview.



*If you are assigned to the post group, I will mail you a packet that has:*

1. A consent form
2. A study survey asking you about yourself (age, gender, education, length of hearing impairment)
3. A survey about how your hearing affects you
4. The SSQ12 survey
5. A release of information so I can get a copy of your hearing test
6. A postage-paid return envelope

*If you are assigned to the interview group, I will schedule a time for you to come to the UC campus for the interview. I will also send you a packet that has:*

1. A consent form
2. A map with directions to the UC campus
3. A \$10 petrol voucher form to help cover the cost of travel to UC

During the interview, I will ask you the same questions as the survey:

1. About yourself (age, gender, education, length of hearing impairment)
2. About how your hearing affects you
3. The SSQ12 survey

**What are my rights?**

You do not have to take part in the study – it is entirely up to you. You can withdraw from the study at any time, without giving a reason. This will NOT affect any future interactions you have with UC. If you do withdraw, I will remove all information relating to you, as long as you let me know by 1 December 2016. After that date, I will not be able to remove your information because it will not be practical.

**What are the benefits of the study?**

There are no direct benefits to you. But, I hope this study will help us provide better support for people with hearing problems by ensuring the surveys we use in clinic are reliable.

**What are the risks of the study?**

There are no direct risks for you being in this study. But, you may feel distressed talking about your hearing problems. You may have whanau or a friend present to help you deal with any distress. You will also find a list of support services at the bottom of this letter.

**Will my information stay private?**

The results of the study may be published, but your identity will be kept private throughout the study. No information that could identify you will be used in any reports in the study. I will not share your information with anyone else.

I will keep the data in a locked filing cabinet and in a password-protected computer. I will destroy the data ten years after I finish the study.

**How can I find out about the study findings?**

Please tick the box on the consent form if you want to know the study results.

**Has this study been approved?**

The study has been checked and approved by the University of Canterbury Human Ethics Committee. If you have a problem or complaint about this research, contact: The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (humanethics@canterbury.ac.nz (03) 364 2987 ext 45588).

**What do I do next?**

If you want to take part in this study, or learn more about it, please contact Rebecca by phone (03 364 2987 ext 8327) or email (Rebecca.kelly@canterbury.ac.nz).

Thank you for taking time to read about this study.

**Who can I contact if I feel distressed?**

Lifeline: 0800 543 354

**Who can I contact if I want more information about hearing loss and hearing aids?**

New Zealand Audiological Society: 0800 625 166

Ministry of Health Healthline: 0800 611 116

Ministry of Health Disability Support: 0800 373 664

## CONSENT FORM

Study title: Test-retest reliability of the *Speech, Spatial, and Qualities of Hearing Scale – Short Form (SSQ12)*: Comparison of administration method and hearing aid experience

The information about this research study has been explained to me to my satisfaction. I have had the chance to ask questions.

I know what I need to do to take part in the study.

I know that I can choose whether or not I take part in this research. I know that I may withdraw from the study until 1 December 2016, without penalty. After that date, withdrawal of information will no longer be feasible. If I withdraw, my information will also be withdrawn.

I know that any information or opinions I give will be kept private to the researcher. I know that any published or reported results will not identify me.

I know that all data collected for the study will be kept in locked and secure facilities or in password protected computers and will be destroyed after ten years.

I will be given a copy of this form and the Research Information Sheet.

I know that I can contact Dr Rebecca Kelly-Campbell for more information. Her contact details are:  
Dr Rebecca Kelly-Campbell: rebecca.kelly@canterbury.ac.nz, (03) 364 2987 ext 8327

If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch ([human-ethics@canterbury.ac.nz](mailto:human-ethics@canterbury.ac.nz), (03) 364 2987 ext 45588).

I would like a copy of the final results of the study.

Yes ☐ No ☐

By signing below, I agree to take part in this research project.

Name (please print): \_\_\_\_\_

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

## Appendix D: Demographic questionnaire.



### Participant Information

*Please answer each question honestly and to the best of your ability*

Date: \_\_\_\_\_ Current age: \_\_\_\_\_ Gender: \_\_\_\_\_

1. What ethnic group do you belong to (please tick all that apply to you)?

- |   |                                  |
|---|----------------------------------|
| <input type="checkbox"/> New Zealand European   | <input type="checkbox"/> Tongan  |
| <input type="checkbox"/> Maori  | <input type="checkbox"/> Niuean  |
| <input type="checkbox"/> Samoan   | <input type="checkbox"/> Chinese |
| <input type="checkbox"/> Cook Island Maori  | <input type="checkbox"/> Indian  |
| <input type="checkbox"/> Other, such as Dutch, Japanese, Tokelauan. Please state: _____ |                                  |

2. What is your relationship status? (please tick one box)

- |                                    |  |
|------------------------------------|--|
| <input type="checkbox"/> Single    | <input type="checkbox"/> Never married               |
| <input type="checkbox"/> Married   | <input type="checkbox"/> In a committed relationship |
| <input type="checkbox"/> Widowed   | <input type="checkbox"/> Divorced                    |
| <input type="checkbox"/> Separated |  |

3. What is the net annual income of your household? (please tick one box)

- |  |   |
|--|---|
| <input type="checkbox"/> \$0 – \$25,000      | <input type="checkbox"/> \$25,000 - \$50,000  |
| <input type="checkbox"/> \$50,000 - \$75,000 | <input type="checkbox"/> \$75,000 - \$100,000 |
| <input type="checkbox"/> more than \$100,000 |   |

4. What is the highest level of education you completed? \_\_\_\_\_

5. On a scale of 1 to 10, how would you describe the severity of your hearing problem (1 = not at all severe, 10 = very severe)? \_\_\_\_\_